A Tourist Place Recommendation and Recognition System - A Plagiarism Report

by Viken Parikh

Submission date: 02-Jun-2018 01:16AM (UTC+0530)
Submission ID: 971308437
File name: Blackbook_B_1.pdf (6.02M)
Word count: 12675
Character count: 66716

Chapter 1: INTRODUCTION

This chapter introduces our final year project, and covers the background, motivation, scope, description and organization of our project report.

1.1 Background

The Travel and tourism industry was among one of the largest industries in the global economy that contributed to over 7.6 trillion U.S. dollars in the year of 2016. Thus for any country, tourism is one of the most impactful sectors and with the advancement of technology, a majority of this tourism has become web-based.

Moreover, users today, different types of tourist places like pictures of landmarks, historic monuments, restaurants or even hotels for stay. Among this, most people do not travel alone. Based on different group preferences (in addition to individual user preferences) have to be modeled, and possibly conflicting requirements must be taken into account to recommend places to the users. People often browse the Internet to express their opinions, give reviews etc. This causes a lot of data available online. Therefore we can leverage this data for making travel recommendations.

The emergence of various machine learning and data mining algorithms has benefitted every industry in some way or the other. However, lots of information can lead to an information overload on the Internet. Moreover, a recommender technology in the tourism field is scarcely applied and people have relied on their acquaintances or tourist guides to suggest places to them which might not be as per their interests. Therefore we aim to leverage various machine learning algorithms to recommend places of interests to a user based on their interests and their similarity to other users.

The field of machine learning and deep learning have made tremendous progress in the last few years on addressing these difficult problems. Specifically, a deep convolutional neural network

can achieve reasonable performance on hard visual recognition tasks like matching, also exceeding human performance in some domains. Therefore using deep learning we have trained a system which can recognize a monument or a tourist attraction. This can contribute immensely to the tourist industry as a user can simply take a photo of any attraction and irrespective of whether it is famous or no, the tourist can get information relevant to the attraction.
Therefore we aim at creating an application that will contribute in the field of tourism.

1.2 Motivation

The prime reason for choosing this topic was our personal interest in traveling and tourism. No matter what the reason is, everyone likes to travel and thus making our Final Year Project in this domain seemed appropriate.

Tourists these days are interested in the more unconventional places, and a visit to these places takes a lot of planning. Initially, this planning was done by travel agents or travel websites, where a customer could choose their plan from various packages. For example, a particular travel company would plan the tours of a particular place, showing the prime tourist attractions. For instance, a company planning a trip to Mumbai (India) would suggest the more common places like Gateway of India or Marine Drive. However, this trip won't satisfy everyone. Therefore, places of interest vary from user to user. This calls for the necessity of a personalized recommendation system, which would recommend places to the users based on their travel interests. We do so by clustering similar users. Therefore users in the same cluster have similar interests, therefore, have a high likelihood of visiting or having interest in the same place.

65
1.3 Scope of the Project

The aim of this project is to provide recommendations for tourists visiting various places and most importantly, suggesting them such places, which they are interested to visit. Additionally, the application also recognizes various places of interest in a city, if it is present in the database.
Functional Requirement

1. An Application for tourist.
2. Suggest places to visit in a city(Hotels, Restaurants, Attractions).
3. Recognise places for which it has been trained.
4. Send and Retrieve data from the server.

Non-Functional Requirements

1. Efficiency – Application will provide quick information.
2. Reliability – Correct location and it's relevant places will be provided.
3. Usability– Coherent and flexible user interface that can be used easily.
4. Security – Only authorized users can modify the system. Only registered users will be able to login and get recommendations.

1.4 Project Description

Our project is divided into two modules i.e the place recommendation module and the place recognition module. The dataset we have used from TripAdvisor which has data for over 1000 users and their reviews for over 20000 places. We first analyzed and pre-processed the dataset.

For the place recommendation module, we use K-Modes as a clustering algorithm for categorical data. Based on the clusters and similarity between the users we make recommendations for places in a given city.

For the image recognition module, we used web scraping to scrap the first 100 images for 55 places in Mumbai. For image recognition, we have used Convolutional Neural Networks. We get an accuracy of 46.4 % for the places our model is trained.

1.5 Organization of the Report

The report on Tourist Place Recommendation and Recognition is arranged in an orderly manner which includes Literature survey as Chapter 2, followed by the Project Design as Chapter 3 which includes the approach and fundamental concepts, Chapter 4 gives the details about the implementation of the project i.e. Recommendation Module and Recognition Module and also includes the screenshots of the implementation. Lastly, the Conclusion states the importance of the project which is followed by the references used and future scope of this project.

Chapter 2: LITERATURE SURVEY

In this chapter we give a background of the related work done till date. Additionally, we also present a few important research papers studied before taking up and during the project.

2.1 Background

Below is a comparative study of various systems we have studied.

	Google Goggles	CamFind	Our Proposed System
Input	Image	Image	Image, Text
Output	i) Recognise image ii) Give Name iii) Web Search Results	i) Recognise image ii) Give Colour iii) Give Name	i) Recognise image ii) Suggest nearby places in the given city based on users' interests.
Image Category	Any Image	Any Image	Images of 1. Hotels 2. Attractions 3. Restaurants

Table 1: Comparison of various similar systems for place recognition.

A Recommendation System aims to provide a list of personalized items that best fit their individual taste. This system analyses the existing data and gives different suggestions to the user on basis of their interests. Greater the understanding user's interest, the better recommendations, it can perform. A few methods for making recommendations are Content-based Collaborative Filtering, Item-based Collaborative Filtering, and knowledge-based filtering. Taking feedback from the user about the recommendation is also a necessary condition for a recommendation system.

2.2 Technical Papers Studied

1. 'You will like it!' using open data to predict tourists' response to a tourist attraction

This paper helped us understand that user-generated content spread via social networking services such as reviews, comments, and past experiences, has been increasing and it has made a great deal of information available. The open data available online in the form of tweets, comments, reviews can play a crucial role to make data-driven decisions from open data. Since users always check about reviews online they play a crucial role in the process of decision making. This paper also helped us find the TripAdvisor dataset which was available online which we are using for our recommendation system.

2. Mobile travel guide using image recognition and GPS/Geotagging: A smart way to travel

In the conventional tourism system, tourists hire guides to know the tourist spot better. But with the elevation in the technology, various applications and services have come up to do this job of the guide. The main function of this paper was to recognize a monument from the picture clicked/uploaded by the tourist, in which the history related to this place. This paper helped us to understand the existing systems and helped us immensely in designing our system

3. Intelligent Search in E-Tourism Services Using Recommendation System: Perfect Guide for Tourist

The author initially discusses the shortcomings of the traditional approaches for recommending the items to the users. These approaches include content-based filtering, collaborative-based filtering and knowledge-based filtering. The author has also proposed three algorithms on recommendation system; one is based on the existing user, second is based on cookies, which are for new users and the last is based on time schedule, with prime importance given to the last one. In Time Scheduling recommendation system a concept of scheduling of activities of the tourist is explored. This provides the real agenda of activities which not only shows the user preferences but also provides the whole schedule which includes how and when to perform activities.

4. User-based Collaborative Filtering for Tourist Attraction Recommendation

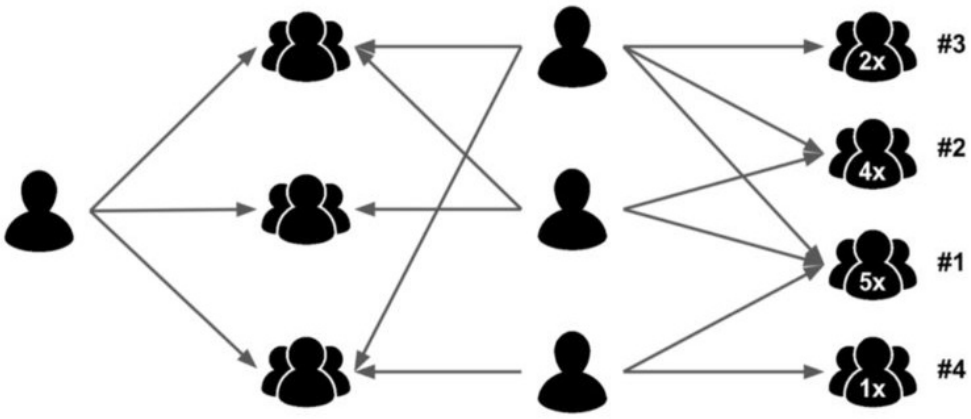


Fig 2.1: User-based Collaborative Filtering (Source: Neo4j Blog)

A user-based tourist attraction recommender system is developed in this paper. The recommender system is used to generate a personalized list of preference attractions for the tourist. On the basis of collaborative filtering principle, the recommendation process of tourist attractions divided into three steps, information representation, neighbor generation and the generation of attraction recommendations. Cosine method is used to find the similarity between two neighbors. Then, according to the neighbor's visiting history, the places are recommended to the user.

5. Implementation of Training Convolutional Neural Networks

Convolutional Neural Networks (CNN) is a deep neural network which can study concurrently. In this article, the author has given a detailed analysis of the process of CNN algorithm using both forward and backward propagation. This algorithm was then applied to implement the face recognition problem in Java. In addition, the author has theoretically analyzed the maximal speed up and the parallel efficiency of the system, by calculating the individual times for forward and backward propagation.

6. K-modes Clustering Algorithm for Categorical Data

Method	Representative	Measure
k-distributions	Distribution	Product of m-estimates
k-histograms	Distribution	non-matching frequencies
k-modes	Mode	Hamming distance
k-medoids	Medoid	Hamming distance
k-entropies	Distribution	Entropy change

Table 2: A list of methods used for clustering of categorical data classified according to cluster representative type and distance measure. [12]

In K-modes clustering technique we choose an initial cluster center modes as seed, which lead towards that problem clustering results to be regularly reliant on the choice initial cluster center and non-repeatable cluster structure may be obtained. K-Modes clustering is widely used for categorical data and replaces means by modes. The algorithm selects attributes on information gain basis to provide a better result. Experimental results show the proposed technique provides a good accuracy for a broad range of datasets.

Chapter 3: PROJECT DESIGN

This chapter describes our problem statement, along with various UML diagrams which will enhance your understanding about the overall working of our system. We further explain in detail various key concepts used in the project. Further we have discussed on our project plan and System Requirements Specification for the project.

3.1 Introduction

Tourism has become an important sector that can boost a country's economy significantly. That being said, it is also important to note that tourism and traveling is one of the biggest entertainment in today's modern lifestyle. And the most crucial task, to carry this out efficiently is to plan the trip. Conventionally this planning was done by travel agents. For example, a particular travel company would plan the tours of a particular place, showing the prime tourist attractions. For instance, a company planning a trip to Mumbai (India) would suggest the more common places like Gateway of India or Marine Drive. However, this trip won't satisfy everyone. If a user is interested in places like museums and aquariums, then they would have to plan such a trip on their own.

Nowadays, mobile phones are a necessary part of people's life. There is a continuous rise in the number of mobile computing applications centered on people's daily life. One of the areas in which the user can benefit from smartphone applications is tourism & traveling.

We aim at combining the comfort of mobile phones and the need of planning tours by proposing the application of Travigate, wherein, on the basis of similar interests with entries in the training dataset, we recommend places to a user. We use K-Modes as a clustering algorithm for categorical data. K-modes is generally used for large datasets and the dataset we have used for this paper is the TripAdvisor dataset. There are entries for about 7000 users and their about 32000 reviews for over 26000 places.

Further, when a tourist visits a place, they are unaware of the place and usually require a tourist guide to tell them which place it is and provide relevant information. Our system has been

trained to recognize a place based on image input of the place using the list of places from the TripAdvisor dataset.

3.2 Problem Statement

1. The given system recommends places a user can visit in a city namely - hotels, attractions, and restaurants based on the interest of the user. The system will have a client-server architecture. When the user enters a place in the mobile application (Eg- Mumbai), the server retrieves information from a MySQL database and displays it for the user.
2. There is also an upload image feature provided, where a user will upload an image of a tourist attraction for which our system has been trained, the system responds with the name of the place.
3. This system also gives an opportunity to the user to add a particular place which is not present in the existing dataset.

3.3 Block/System Diagrams

3.3.1 System architecture

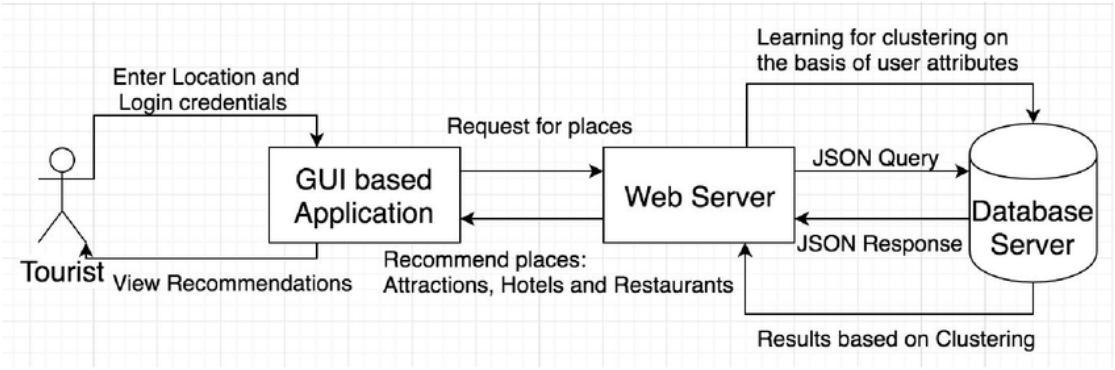


Fig 3.1: System architecture of Place Recommendation System

There are 4 major components in the system architecture:

1. The user
2. The Android Application (Travigate)
3. The Web Server
4. The Database Server

The interactions between the user and the application are through the user interface, where the users insert their credentials and the output is displayed on their screens.

The application then sends a request for the places to the Web server, which in turn fetches data from the Database by sending JSON Queries.

The clustering using K modes is done on this dataset and the results are sent back to the Web server by JSON responses. The web server sends these correctly clustered attractions, hotels, and restaurants to the application. The user views these places as the recommendations.

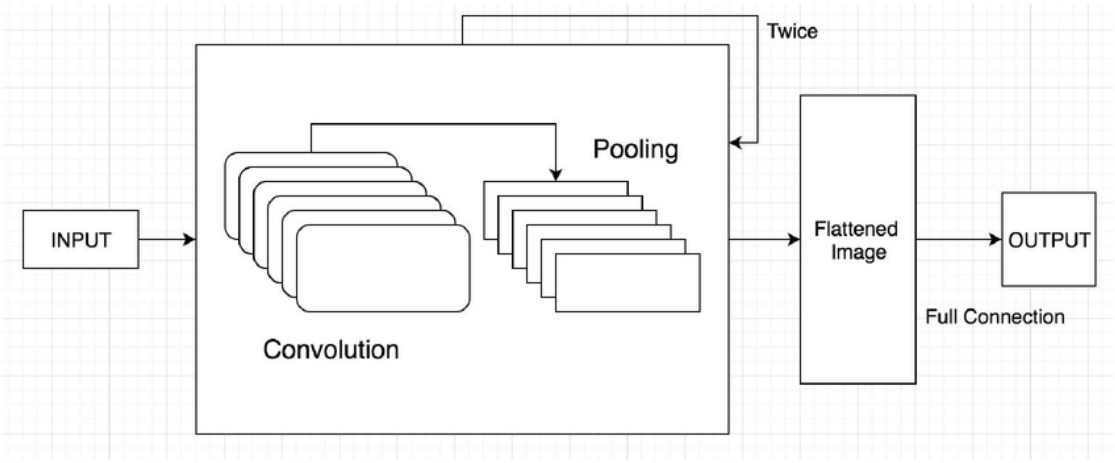
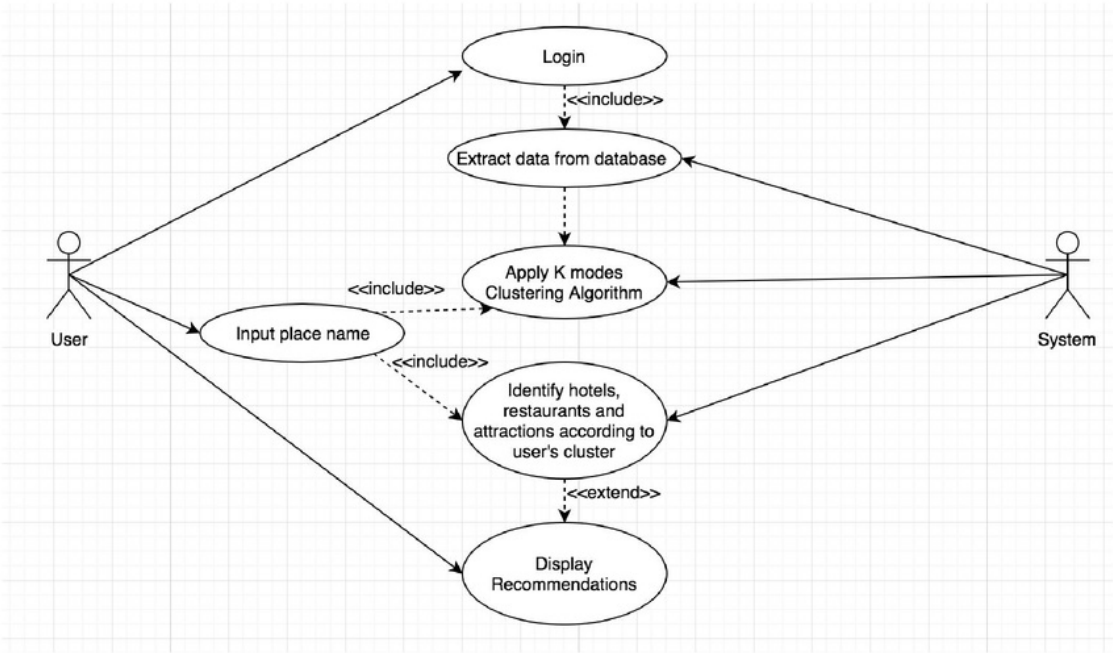


Fig 3.2: System Architecture of Place Recognition System

3.3.2 Use case Diagram



60

Fig 3.3: Use case diagram for Place Recommendation System

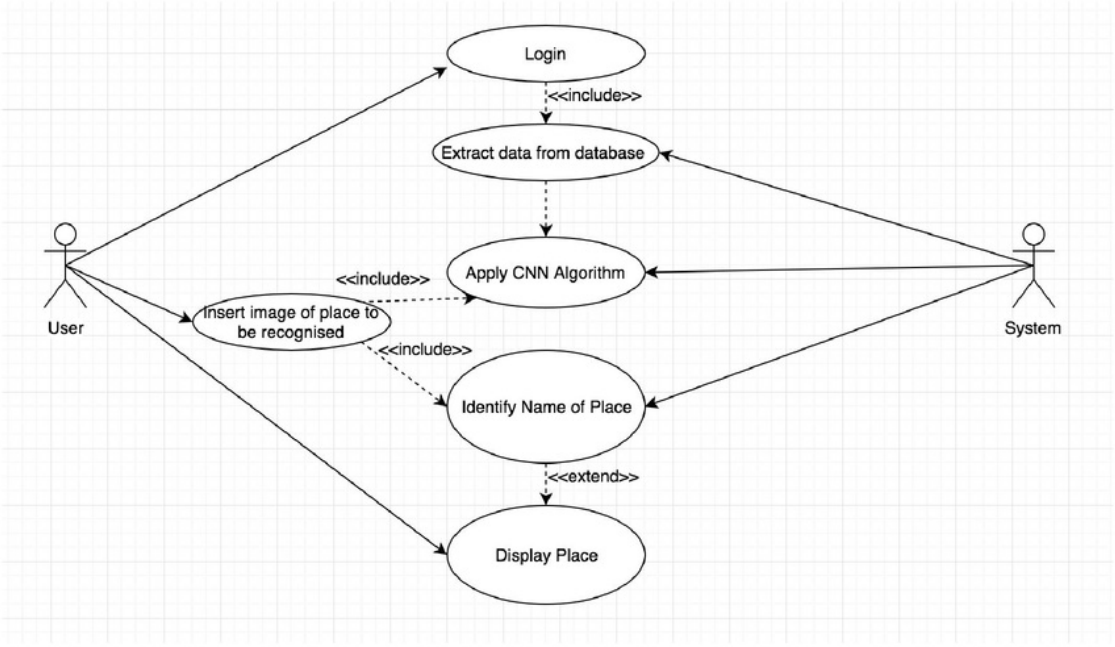


Fig 3.4: Use case diagram for Place Recognition System

3.3.3 Sequence Diagram

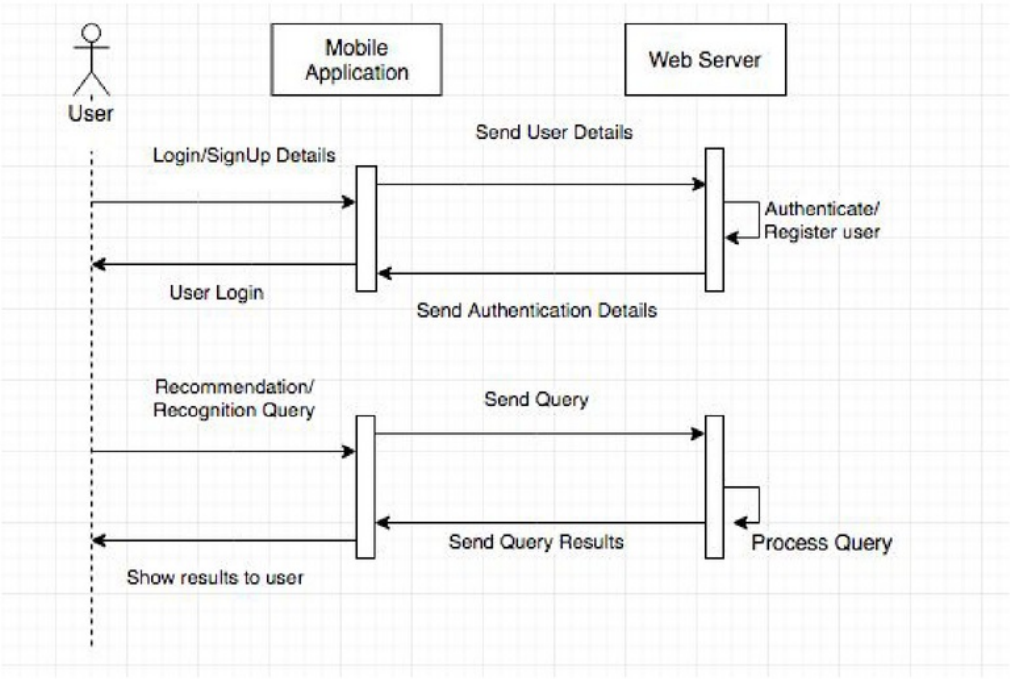


Fig 3.5: Sequence Diagram of the application

3.3.4 Data Flow Diagram

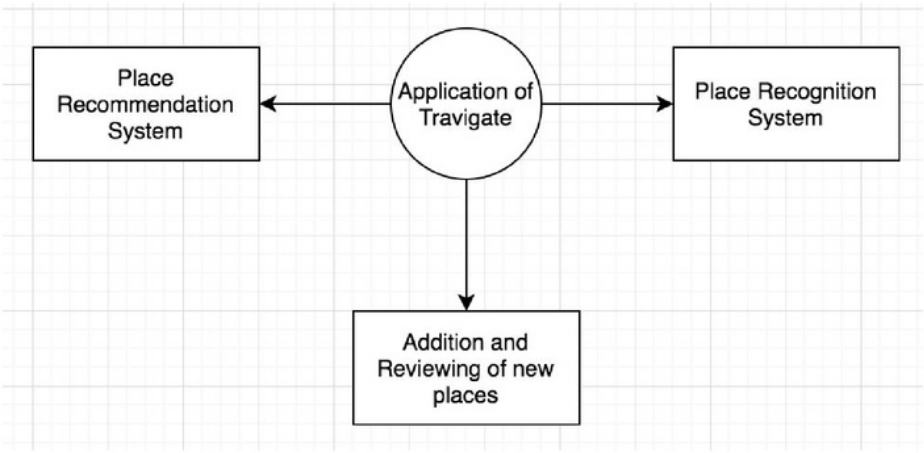


Fig 3.6: Level 0 Data Flow Diagram

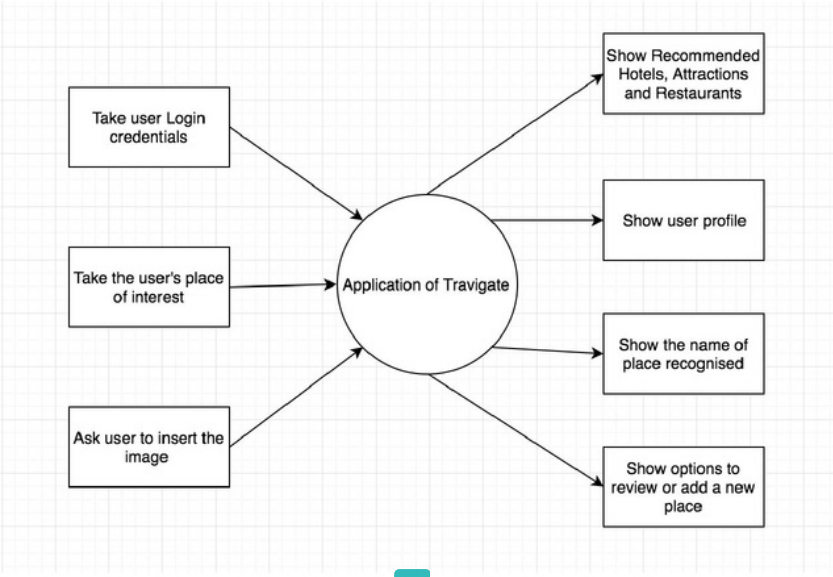


Fig 3.7: Level 1 Data Flow Diagram

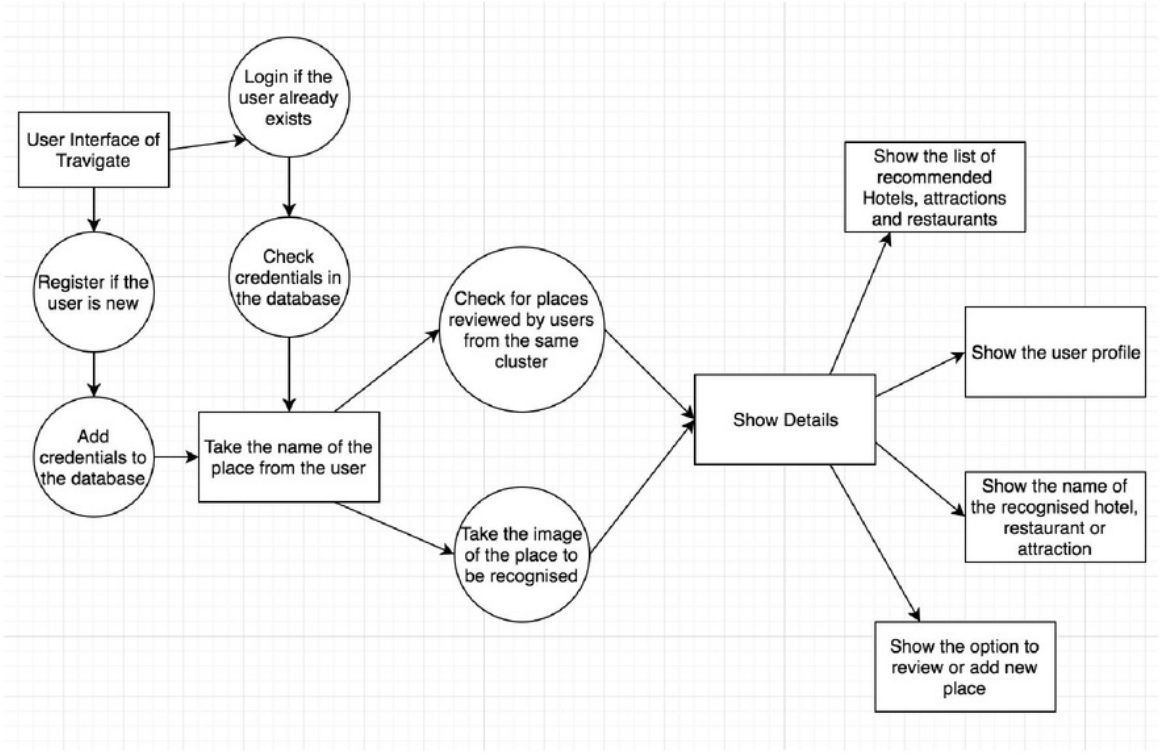


Fig 3.8: Level 2 Data Flow Diagram

3.4 Objective

The primary objective of this project is to create a tourist place recommendation and recognition system which recommends places (Attractions, Hotels, Restaurants) to a user based on their interest. Thereby making traveling more interesting as a user can visit places of their interest rather than certain mainstream places which are visited by most people. The system also intends to recognize a place, when a user uploads a photo of it. Thereby we aim to make traveling easier, more interesting and easily accessible as we deliver our application across mobile platforms.

3.5 Key Concepts in the Project

3.5.1 Clustering and K- Modes Algorithm

Clustering is an unsupervised learning problem; it aims at finding a structure and patterns in a collection of unlabeled data. Clustering can be understood as grouping objects having some measure of similarity together. Therefore by grouping similar objects in clusters, we can analyze the clusters and find various patterns in the data, which can help us label the data. Various clustering algorithms are K-means, Hierarchical clustering, DBSCAN etc.

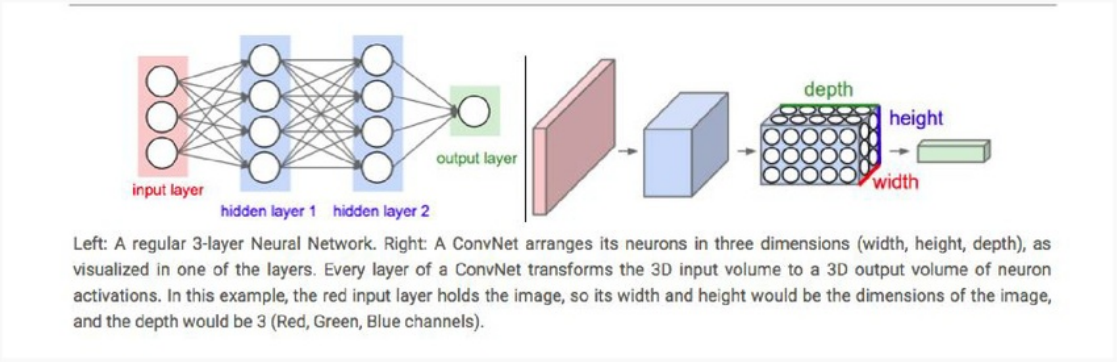
However mainstream, clustering algorithms are not sufficient since most of them only work for numerical data. However even categorical data exists and it is important to cluster data where certain attributes have categories like Male, Female. In K-Means calculate the Euclidean distance as a distance measure based on which we form clusters. However, Categorical data is different. For example, {male, female},{low, medium, high} are examples of categorical data. It is not possible to find the distance between male and female. Traditional clustering techniques fail to handle datasets with categorical attributes since it minimizes the cost function by calculating the means. The K-means based Partitioning clustering methods are used to process large numeric data sets for its simplicity and efficiency. Clustering technique can be generally classified into two groups: hierarchical, partitioning clustering. The k-modes algorithm is an extension to k-means for categorical data, by replacing k-means with k-modes, we introduce a different dissimilarity measure and update the modes with a frequency based method. Instead of distances, k-modes uses dissimilarity i.e smaller the dissimilarity more similar are the objects.

3.5.2 Convolutional Neural Network

Convolutional neural networks are deep artificial neural networks that are used to classify images. Neural Networks take inspiration from the human brain. Humans can see multiple images every second and process, however, we are not aware how the processing is done.

However, machines work differently. A machine will see an image into a matrix of pixels of different brightness and color.

A CNN consists of an input layer, an output layer and multiple hidden layers in the middle. A CNN consists of one to multiple convolutional layers followed by one or more fully connected layers similar to the standard multilayer neural network. The architecture of a CNN is designed to take advantage of the 2D structure of an input image. This followed by the pooling which reduces the number of pixels and dimensionality of the image, thereby increasing computing speed. Moreover, CNNs' are easier to train and have many fewer parameters than fully connected networks with the same number of hidden units.



Left: A regular 3-layer Neural Network. Right: A ConvNet arranges its neurons in three dimensions (width, height, depth), as visualized in one of the layers. Every layer of a ConvNet transforms the 3D input volume to a 3D output volume of neuron activations. In this example, the red input layer holds the image, so its width and height would be the dimensions of the image, and the depth would be 3 (Red, Green, Blue channels).

Fig. 3.9: Comparison between Neural Networks and Convolutional Neural Network(Source: Stanford University)

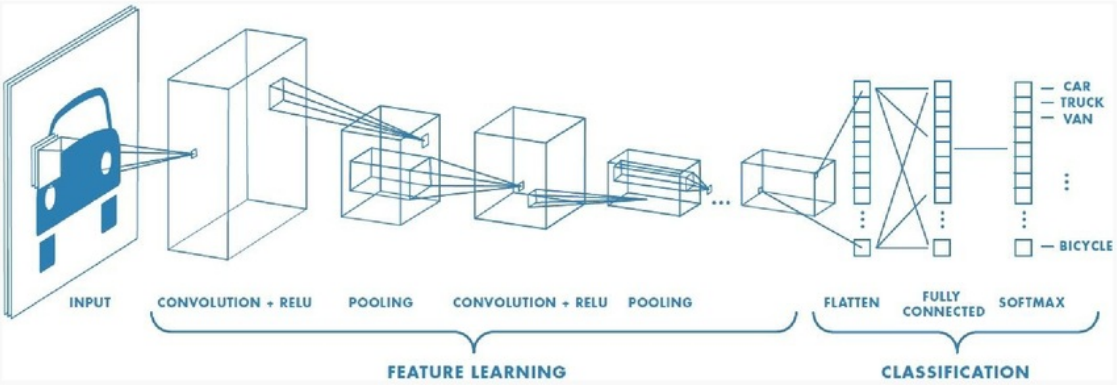


Fig. 3.10: Flow of Convolutional Neural Network(Source: Mathworks)

The main layers in CNN architecture are : Convolutional Layer, Pooling Layer, Activation Layer, Fully-Connected Layer. The layers are explained in detail below:

Convolutional Layer:

This is the first layer of the CNN, it's motive is to extract features from the input image.

The Convolution layer consists of various of learnable filters. Every filter is small spatially compared to the size of the image, but it can extend throughout the full depth and input volume. In this layer, a 2-D convolution is performed on the inputs. During the forward pass, a convolution is performed between the filter and a part of the image. We slide the filter for the entire image.

Images are high dimensional inputs, so it is impossible to connect neurons to all neurons of such a huge volume. Instead, the connection of each neuron is only to a local region of the input volume.

There are 3 hyper-parameters in Convolutional Layer which determine the size of the output volume.

1) The **depth** is the number of filters, we want to use. For convolution, we can use multiple filters for different inputs.
2) The **stride** with which we want to move the filter.
3) **Zero-padding** i.e adding zeros to the border, this enables us to control the size of the output volume.

Pooling layer:

Dimensionality of the image is reduced using the pooling layer, this, in turn, helps in decreasing the number of parameters of the model and it's computational complexity. One of the major advantages of this layer is the increase in the computational speed. The most used form of pooling is called MAX pooling, where the maximum value in the matrix is taken. Other known forms include AVERAGE pooling, in which the average of all the values in the matrix is calculated as the output of pooling.

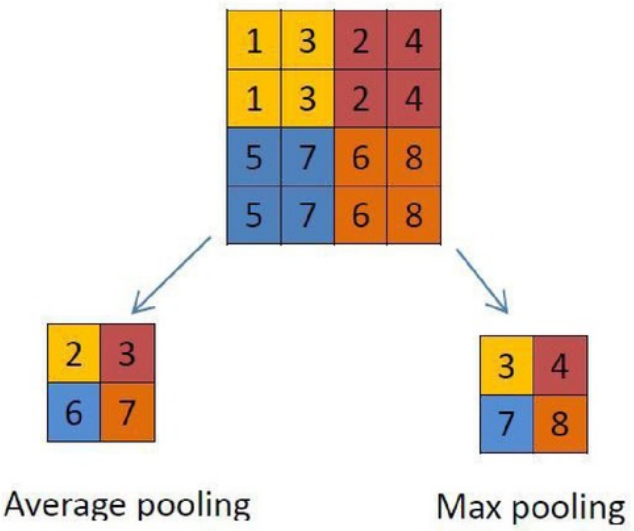


Fig. 3.11: Example of Max and Average Pooling (Image source: towardsdatascience.com)

Fully-connected layer:

Fully connected layers connect every neuron in one layer to neurons in other layers, in CNN this layer is similar to that of the mainstream neural networks. The activations are calculated in the fully connected layer with various activation functions like ReLU, Softmax.

3.5.3 Web Scraping

Web Scraping is extracting data from websites and the data is extracted and saved to a local file in your computer.

Data displayed on most websites can only be viewed using a web browser. Web browsers do not offer the functionality to save a copy of this data for personal use. Therefore, one either manually copy and paste the data - a which is a very tedious and time-consuming approach. Web Scraping basically automates this process, so that instead of manually copying the data from websites, a Web Scraper will perform the same task within a fraction of the time. We use web scraping in our project to scrap the first 100 images for a particular tourist attraction on Google. Through web scraping, we scraped over 5500 images of 55 places in Mumbai which we use to train our CNN model.

3.5.4 JSON

JSON stands for JavaScript Object Notation. JSON is a language-independent text format that uses conventions of many programming languages making it an ideal data-interchange language. It is built on universal data types supported by all modern programming languages making JSON an ideal data-interchange language.

3.5.5 Client-Server Architecture

The client-server architecture is a network architecture with two components, a Client, and a Server. In this type of architecture, there are one or more client computers connected to a central server over a network or internet connection. All the systems using a Client Server Architecture have shared computing resources.

- **Servers** are the process dedicated systems, they do the tasks like managing file servers, network traffic, printers and disk drivers.
- **Clients** are on the user side. Users run applications on these client PCs. Clients have to rely on servers for all their resources like devices, power, file and so on.

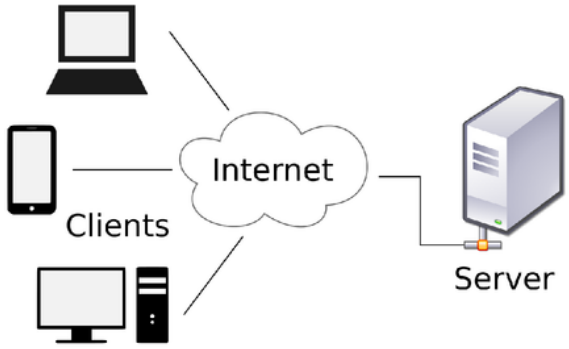


Fig. 3.12: Client Server Architecture (Image source: https://en.wikipedia.org)

3.6 Feasibility Analysis

3.6.1 Economic feasibility:

Economic feasibility states whether the firm can afford to build the software and whether its benefits should substantially exceed its cost. Our project is economically feasible. Our system uses Python as the back end programming language and the Ionic framework for the front end; which was very feasible, economically since it is free of cost.

3.6.2 Technical feasibility:

In technical feasibility, we see whether the technology needed for the system exists and how difficult it is to build. Our project can work on almost all the Android phones, thus improving its technical feasibility. Software used for the project implementation is Spyder (Anaconda). Basic technical knowledge of operating Spyder software along with the knowledge of Python is required for the developers.

3.6.3 Schedule Feasibility:

The project is entirely built from scratch to completion in a span of eight months. Thus, we can say that our project is schedule feasible.

3.6.4 Ecological Feasibility:

Ecological Feasibility helps to decide whether the system has an impact on its environment or not. There are no adverse effects of our project on the environment.

3.6.5 Operational feasibility:

The system is easy to use and user-friendly. All maintenance issues will be handled efficiently. The system is adaptable to most environments. Hence our system is operationally feasible.

3.7 Scheduling and Work Allocation

	Task Name	Duration	Start	Finish	Predecessors
1	Preliminary Tasks	10 days	Tue 8/1/17	Mon 8/14/17	
2	Analysis tasks	40 days	Tue 8/15/17	Mon 10/9/17	1
3	Agreement	10 days	Tue 10/10/17	Mon 10/23/17	2
4	Design tasks	25 days	Tue 10/24/17	Mon 11/27/17	3
5	Risk Analysis	12 days	Tue 11/28/17	Wed 12/13/17	4
6	Implementation	45 days	Thu 12/14/17	Wed 2/14/18	5
7	Testing and Integration	30 days	Thu 2/15/18	Wed 3/28/18	6
8	Deployment	12 days	Thu 3/29/18	Fri 4/13/18	7
9	Support	60 days	Mon 4/16/18	Fri 6/8/18	8

Fig. 3.13: Time Scheduling with start and finish date allocation

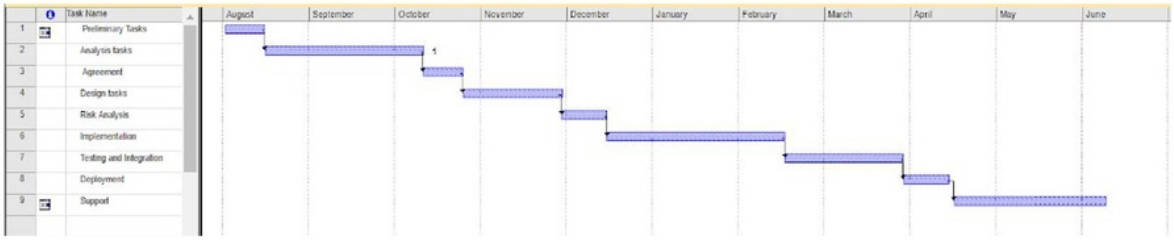


Fig. 3.14: Gantt chart for scheduling task

3.8 Lifecycle Model

Waterfall model is non-iterative design process where System requirements are known initially and the final outcome is determined. It progresses steadily downwards through above-given faces.

16

When to use the waterfall model:

- This model is used only when the requirements are very well known, clear and fixed.
- Product definition is stable.
- Technology is understood.
- There are no ambiguous requirements
- Ample resources with required expertise are available freely

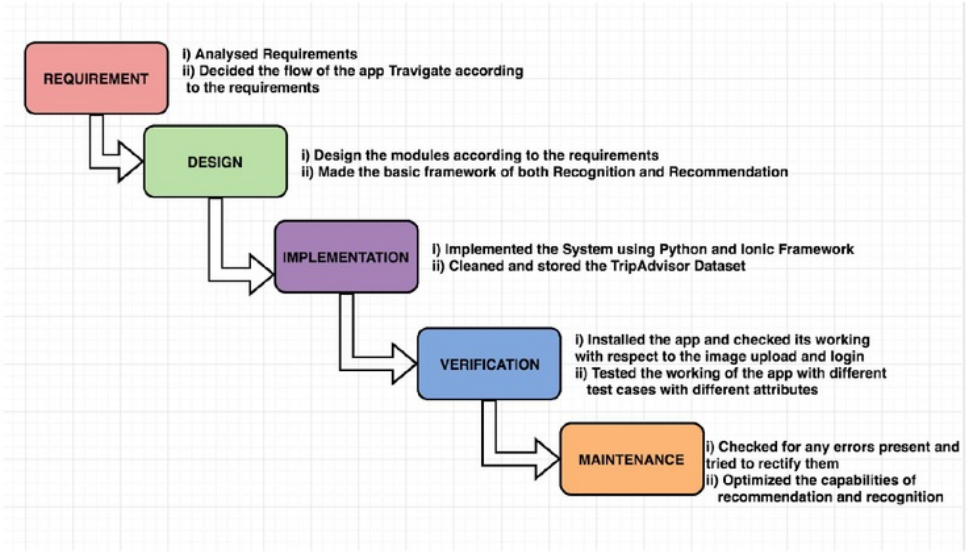


Fig. 3.15: Waterfall model of Travigate system

3.8.1 Waterfall Model Functionality 1: Requirements Gathering

1. Two different datasets have been used. For users' profiles and places' reviews database, the Tripadvisor dataset has been used. For image recognition which consists of images crawled from Google.
2. Clustering algorithms for recommendation module- K Modes
3. Feature extraction techniques and Classifiers such as CNN

Functionality 2: Design

Designing the process overview, from applying clustering to using CNN for image recognition.

Functionality 3: Implementation

1. Implementing all algorithms in Python.
2. Implementing User Interface using the Ionic framework.

Functionality 4: Verification

Verified the system using 20 different users from different clusters from varying locations.

Functionality 5: Maintenance

Maintaining from time to time for its efficiency.

3.8.2 Project deliverables

1. Software Project Management Plan
2. Software Requirements specifications
3. Software Design Description
4. System Test Document
5. User Interface module Prototype
6. User Manual
7. Final Product

3.9 Work Breakdown Structure

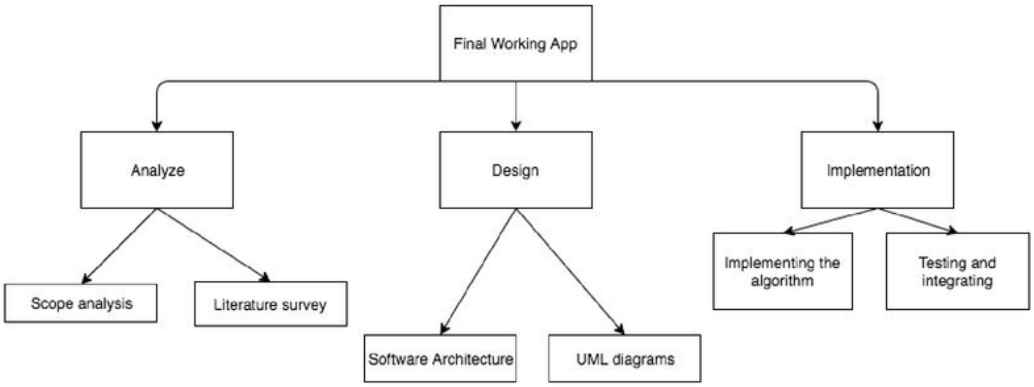


Fig. 3.16: Work Breakdown Structure

Chapter 4 - IMPLEMENTATION AND EXPERIMENTATION

In this chapter we have provided a detailed explanation of the implementation of the two major modules of our project i.e recommendation and recognition modules. We have also given a brief description of the programming languages used, along with screenshots of the application developed.

Our project mainly had two modules to implement, i.e the recommendation module for which we clustered similar users using K-modes and the image recognition module for which CNN was used. Additionally, we also created a mobile application using Ionic Framework and connected it to a server using Flask.

4.1 Dataset Description

Two different datasets have been used. For users' profiles and places' reviews database, the Tripadvisor dataset has been used. For image recognition which consists of images crawled from Google.

4.1.1 TripAdvisor Dataset

There are entries for about 7000 users and their about 32000 reviews for over 26000 places across the world.

a. TripAdvisor User Dataset:

Each reviewer can choose a travel style among 17 different categories (Multiple options allowed) such as Foodie, Luxury Traveler, Shopping Fanatic, Urban Explorer, Peace and Quiet Seeker, Art and Architecture Lover, Thrifty Traveler, Vegetarian, History Buff, Thrill Seeker, Beach Goer, Family Holiday Maker, Nature Lover, Trendsetter, Like a Local, Nightlife Seeker, 60+ Traveler. There are also several other entries such as age range, Number of Helpful reviews etc.

b. TripAdvisor Reviews Dataset:

There are mainly 3 categories of places (Hotels, Restaurants, and Attractions). Each tuple has the following entries- username, type, date, title, text, rating, helpfulness, total-points, Name of Place, City (where the place is located).

4.1.2 Images Dataset

The images dataset has been crawled from Google Images for the places mentioned in the TripAdvisor dataset.

For our training model 55 places that have been reviewed in Mumbai (Bombay) from the TripAdvisor dataset with 100 images for each place which has been achieved using our web crawler designed for scrapping first 100 images from google images search. We split this data into 70 images training, 20 images validation and 10 images for testing each class in the dataset.

4.2 Implementation Steps for Place Recommendation System

The system consists of a MySQL database, a Web Server, and a front-end application.

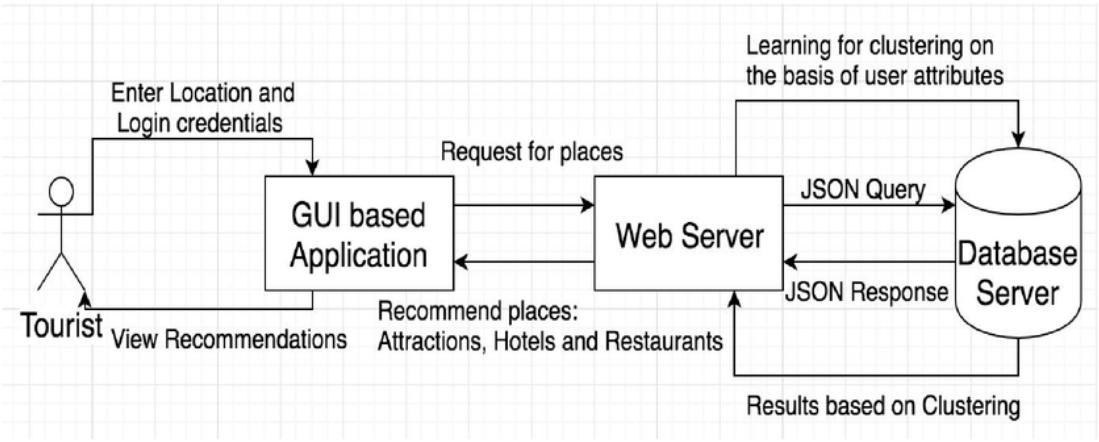


Fig. 4.1: Working of Place Recommendation System

A. Data Cleaning and Data Pre-Processing:

The initial step the data was initially in the form of .csv files it had to be converted into a MySQL database. The dataset also required cleaning and removing redundant or missing data from the datasets. Moreover, it was also necessary converting all the data into categorical data as we have used the modes algorithm which clusters categorical data. The data was loaded to MySQL using a python script.

B. Recommendation Using K-Modes:

The users are clustered using K-modes clustering algorithm, where the total number of clusters can be n (n>2). As the number of clusters increases, the number of users in each cluster decrease. Based on the clusters, if a user wants recommendations for places in a particular city, the system would recommend places users have visited in the same cluster.

For example, if the user wants recommendations, for places to visit in Mumbai (Bombay). The system would recommend places in Mumbai (Bombay) which have been visited by users in the same cluster.

$$d(X, Y) = \sum_{j=1}^{m} \delta(x_j, y_j) \qquad (1)$$

where,

$$\delta(x_j, y_j) = 0, \quad x_j = y_j \qquad (2)$$

$$= 1, \quad x_j \neq y_j$$

Where d(X, Y) is the distance function, to find the dissimilarity between X and Y described by m categorical values.

C. Place recommendation (K-Modes) code implementation steps:

Python 3.6 and its libraries have been used for implementation.

1. Import required libraries like pandas, numpy and others for data manipulation
2. Import the data from MySQL database to a python dataframe using pymysql.
3. Convert Travestyle from string attributes to binary/single-valued attributes.
4. Merge data from users and reviews dataframes.
5. Divide the data for reviews from
6. Create a Kmodes model using KModes library (library source : https://github.com/nicodv/kmodes). Select the number of clusters as 4 and initialize centroids according to the method by Huang [1997].
7. Fit the data to the model and save them to an SAV file (which is a file extension for : Statistical Package for the Social Sciences)
8. While recommending places to a user, the trained model is first loaded.

9. When a user data query is sent from the client side, the cluster of the user is predicted, and based on the reviews of other users' in the same cluster, the places are shown to users as recommended and not recommended.

4.3 Implementation Steps for Place Recognition System

A. Image Scraping:

To create a sufficiently large training dataset for different places, we used the web scraping technique, wherein the first hundred images of a Google search of that particular place are scraped.

In the image search page created by Google, the first hundred websites are visited and the images from those websites are downloaded and stored in individual folders.

This method is convenient when we do not have an existing dataset. We have attempted to use this technique to acquire many photos of a particular place, in order to train the system.

B. Image Scraping module code implementation steps:

Python 3.6 and its libraries have been used for implementation.

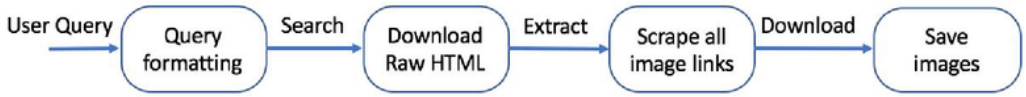


Fig. 4.2: Workflow for scraping images from google (Source: github.com/hardikvasa/google-images-download)

1. Select a city that user wants to train for image recognition.
2. From the reviews data from MySQL, select and load the list of unique places that users have posted their reviews about in the dataset to a dataframe
3. For each place, we set a limit of 100 images to save from google.
4. We create a Google query for each place keyword.
5. Then we download the raw Html of this results page.
6. The folder of the place is created.
7. All the image links are extracted from this raw HTML page and the images are all saved in the folder with the name of the place.
8. These steps are repeated for each unique place in the city that is there on our list.

C. Initializing CNN:

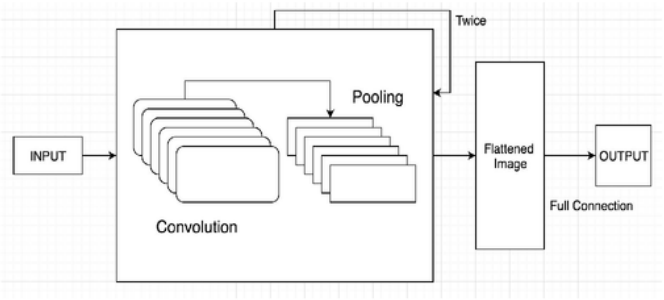


Fig. 4.3: Working of Place Recommendation System

There are a number of important choices that need to be made when we are building and training a neural network. We had to decide how many layers to have, what stride and kernel size to use for each convolution layer and many more. But amongst all these, weight initialization has the greatest impact on both the convergence rate and final quality of a network.

Thus, in this part of the implementation, we initialize the weights used in this Convolutional Neural Networks. For the convolution layer, we have taken the size of the input matrix as 64*64 and apply a 3*3 convolution filter.

D. Convolution:

After procuring these images and initializing the weights, the next step is the application of the convolution layer.

Convolution is basically a mathematical operation which provides a way of multiplying two arrays of numbers, generally of different sizes, but of the same dimensionality, to produce a third array of numbers of the same dimensionality.

In our code, we have used the method Conv2D(), which is using a 2D convolution layer, this layer creates a convolution kernel that is convolved with the layer input to produce a tensor of outputs.

The size of the input of convolution matrix we have used is 64*64 and 3*3 convolution filter. For activation we have used the Relu function.

E. Pooling:

A pooling layer is periodically inserted in between two convolution layers. This layer, is used here to downsample the volume spatially, independently in each depth slice. The method we have used is that of MaxPooling2D(). This is used for the maximum pooling, between two convolution layers.

F. Flattening:

After another convolution layer and a max pooling layer which reduces height and width of the individual channels in favor of their number, i.e., depth, we use the flattening layer.

For the application of the last layer, the dense layer, it is necessary to convert the output of the convolutional part into a 1D array. This is achieved by flattening the output of the initial layers.

G. Dense Layer:

The dense layer is the third layer in the Convolutional Neural Networks.
In a dense layer, all the nodes are connected i.e., every node in the layer is connected to every node in the preceding layer. A dense layer thus is used to change the dimensions of your vector. For multiple output classes, softmax activation function gives the best results.

H. Convolutional Neural Network module code implementation steps:

Python 3.6 and its libraries have been used for implementation.

1. Initialise a sequential model.
2. Add a 2-dimensional convolutional layer with relu activation function and 3*3 convolutional filter to this sequential model.
3. Next, add a 2-dimensional Max pooling layer of a 2*2 size to this sequential model.
4. Add another 2-dimensional convolutional layer with relu activation function and 3*3 convolutional filter to this sequential model.
5. Add another 2-dimensional Max pooling layer of a 2*2 size to this sequential model.

6. Add another layer to flatten the 2 dimensional to 1 dimensional
   Add last 2 dense layers, first with 128 units and relu activation function and other with units= number of output classes and softmax activation function.
7. Finally, compile this model using adam optimizer.
8. Then load the dataset of places images downloaded using web scraping.
9. Fit the sequential model using this dataset.
10. Then save the model in a JSON format and its weights in an H5 format.
11. When the user wants to recognize a place, this model is loaded and will be used to predict an output class of the image.

4.4 Programming language used for implementation

1. **Python**

   Python is a high-level language which is used in general programming. It features a dynamic system. It has multiple programming paradigms, including object-oriented, functional and procedural, and also consists of a large and comprehensive standard library.

   Python focuses on code readability, which means that Python helps the coder to code in fewer steps as compared to other high-level languages like Java or C++.

   We have used Python in our project, because of its versatile features and fewer programming codes. Today, approximately 14% of the programmers use Python on their operating systems like UNIX, Linux, Windows and Mac OS.

2. **Ionic Framework**

   Ionic is a complete open-source SDK for hybrid mobile app development. It is built on top of AngularJS and Apache Cordova. The more recent releases, known as Ionic 3 or simply "Ionic", are built on Angular.

   We have chosen the Ionic framework to build our app because of its numerous features to enhance the user interface like forms, filters, action sheets, navigation menu, list views, tab bars and other UI paradigms.

4.4 Implementation Screenshots

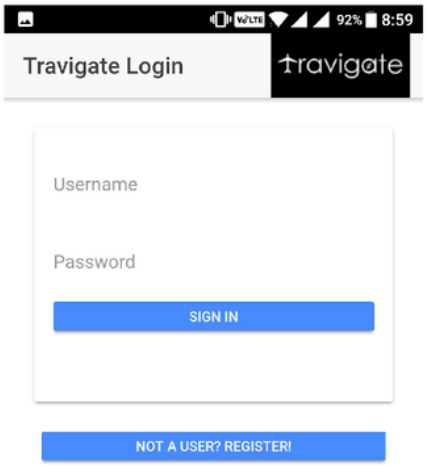


Fig. 4.4 (a): The Login page of the Travigate application.

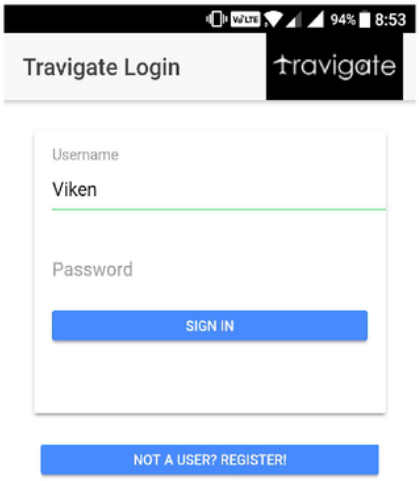


Fig. 4.4 (b): Existing users has to fill in username and password

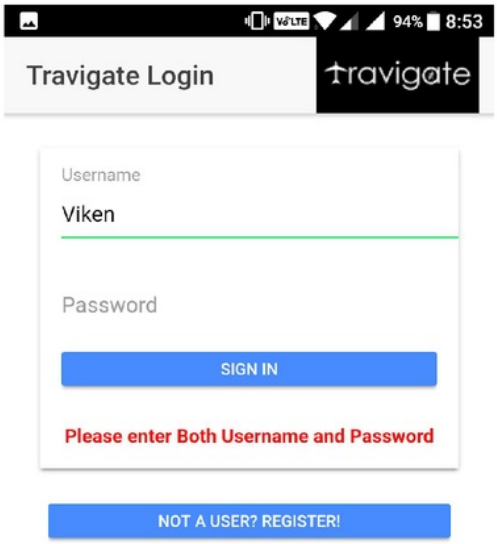


Fig. 4.4 (c): The login will not be considered if all the fields are not filled

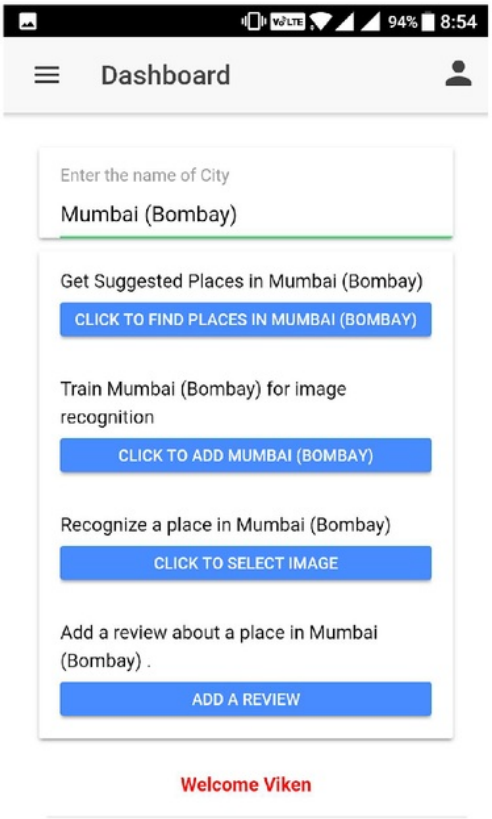


Fig. 4.4 (d): The main dashboard with all the options available.

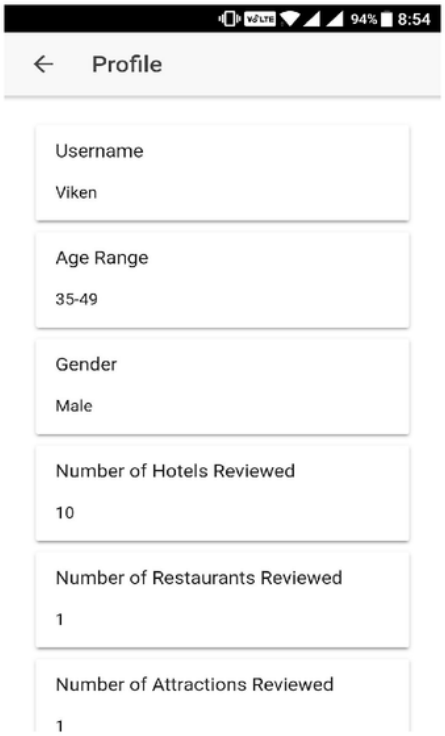


Fig. 4.4 (e): All the attributes of the user are shown in his profile

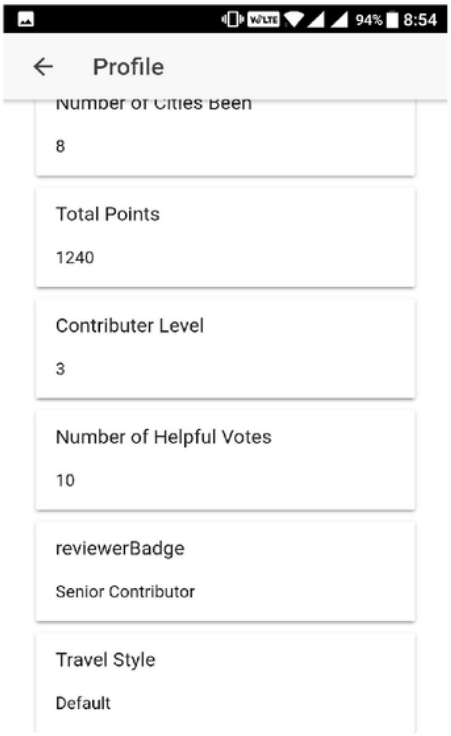


Fig. 4.4 (f): All the attributes of the user are shown in his profile

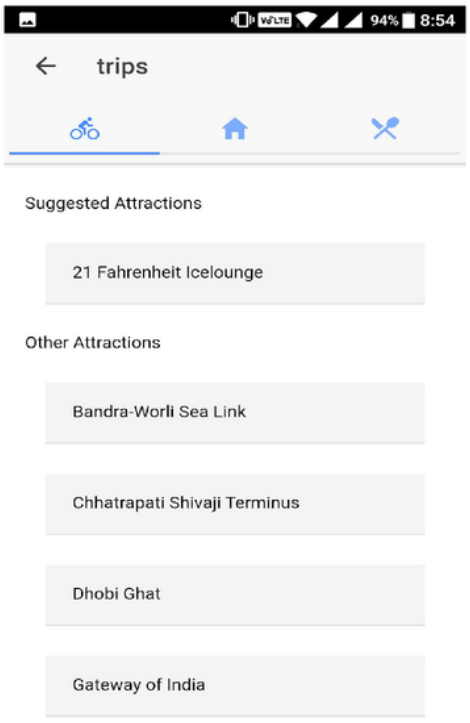


Fig. 4.4 (g): Recommended attractions to the user in Mumbai(Bombay)

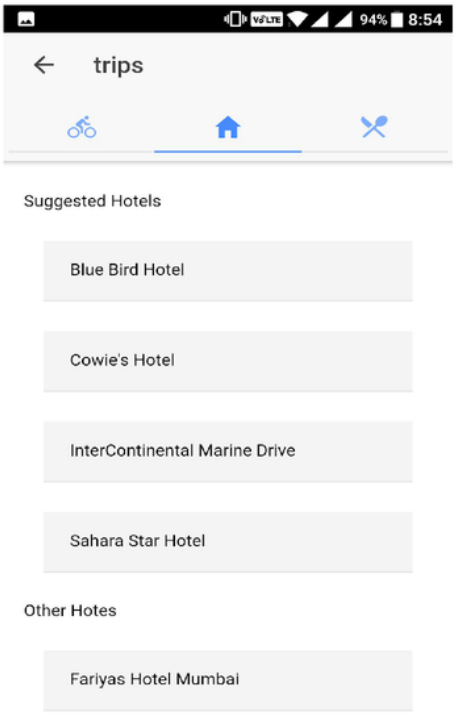


Fig. 4.4 (h): Recommended hotels to the user in Mumbai(Bombay)

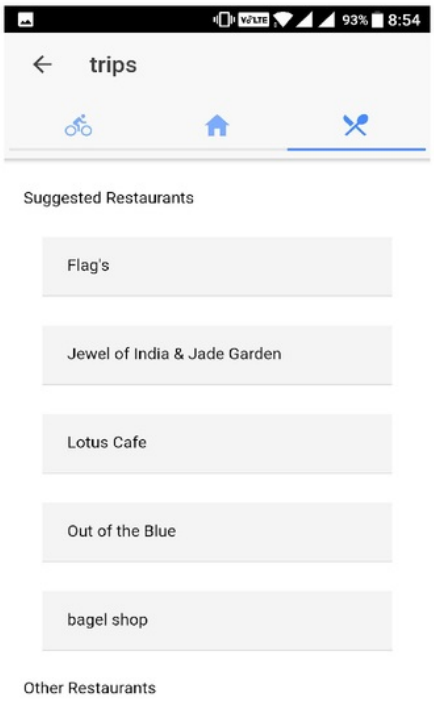


Fig. 4.4 (i): Recommended hotels to the the user in Mumbai(Bombay)

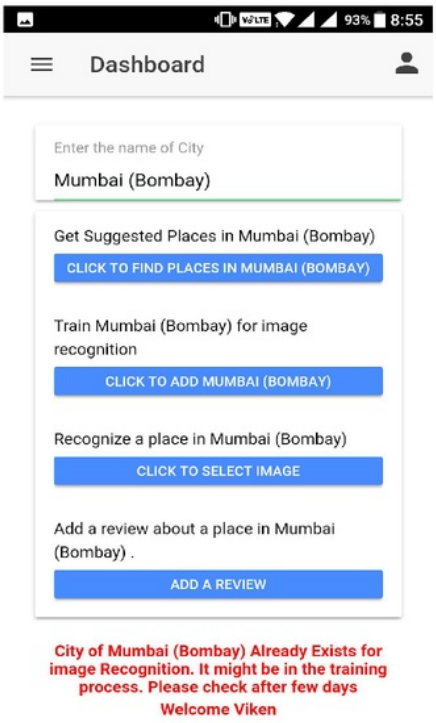


Fig. 4.4 (j): New places can be added to train the system, existing places will not be taken

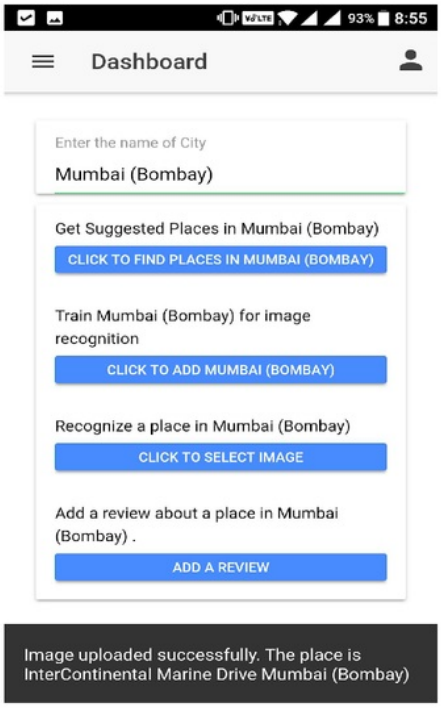


Fig. 4.4 (j): Place recognition on uploading of an image.

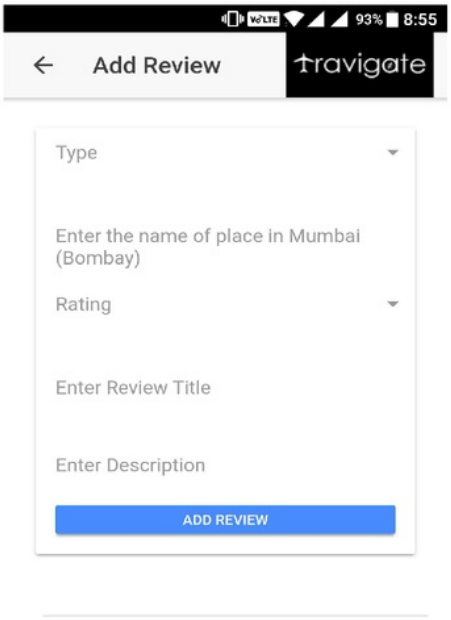


Fig. 4.4 (k): The app allows to review any place from the dataset.

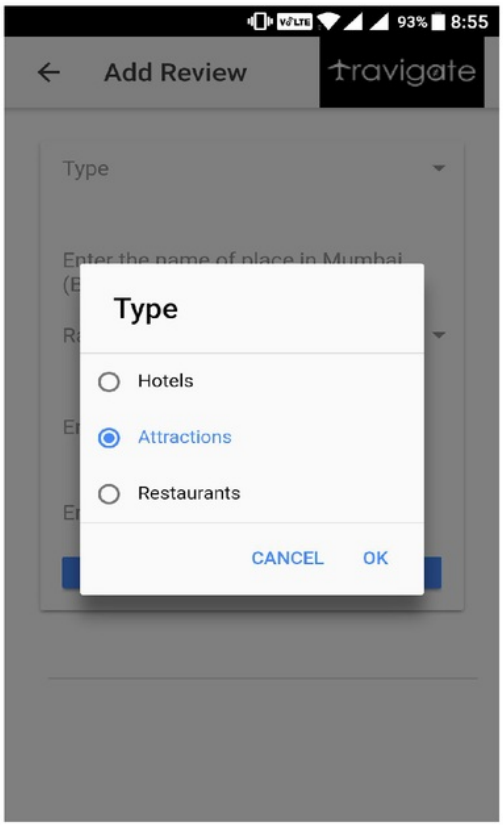


Fig. 4.4 (l): The user has to select which place he is going to review

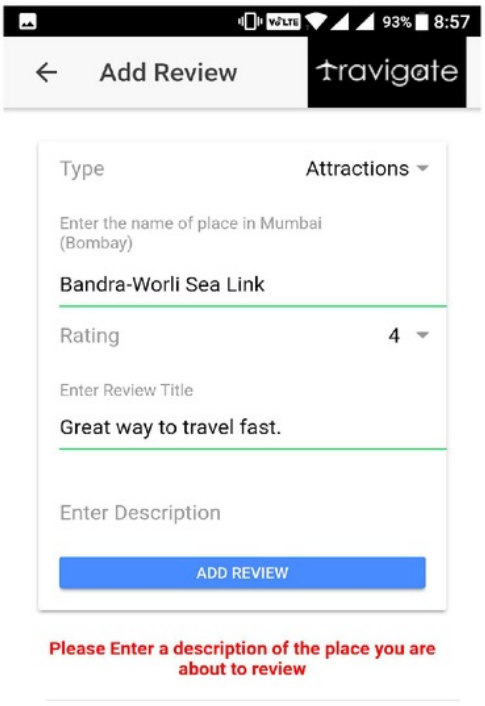


Fig. 4.4 (m): Details of the place, along with rating is to be mentioned

74
Fig. 4.4: Application Screenshots

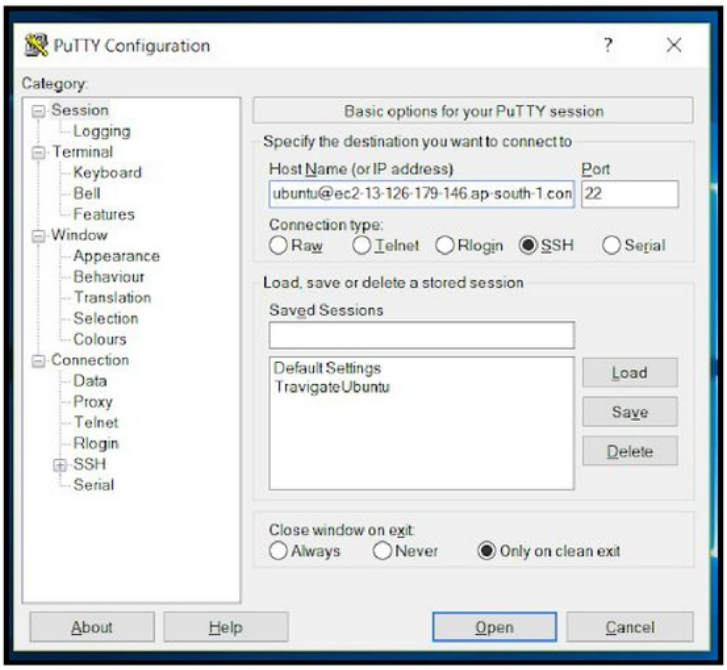


Fig. 4.5: PuTTy configuration

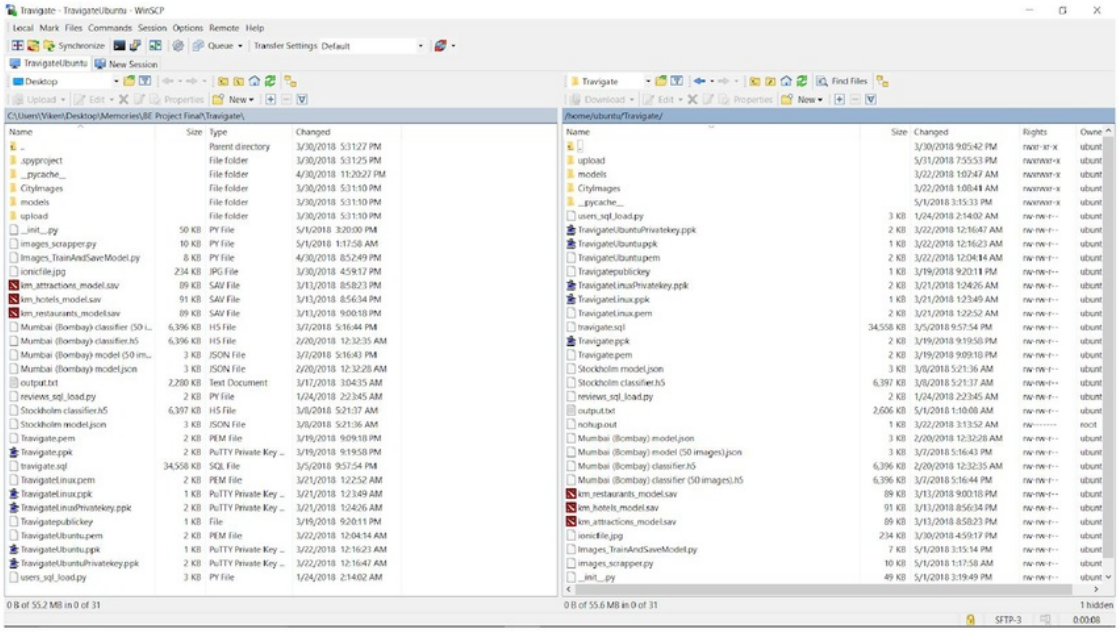


Fig. 4.6: Uploading the python scripts from the local Microsoft Windows machine to Amazon Elastic Compute 2 using WinSCP.

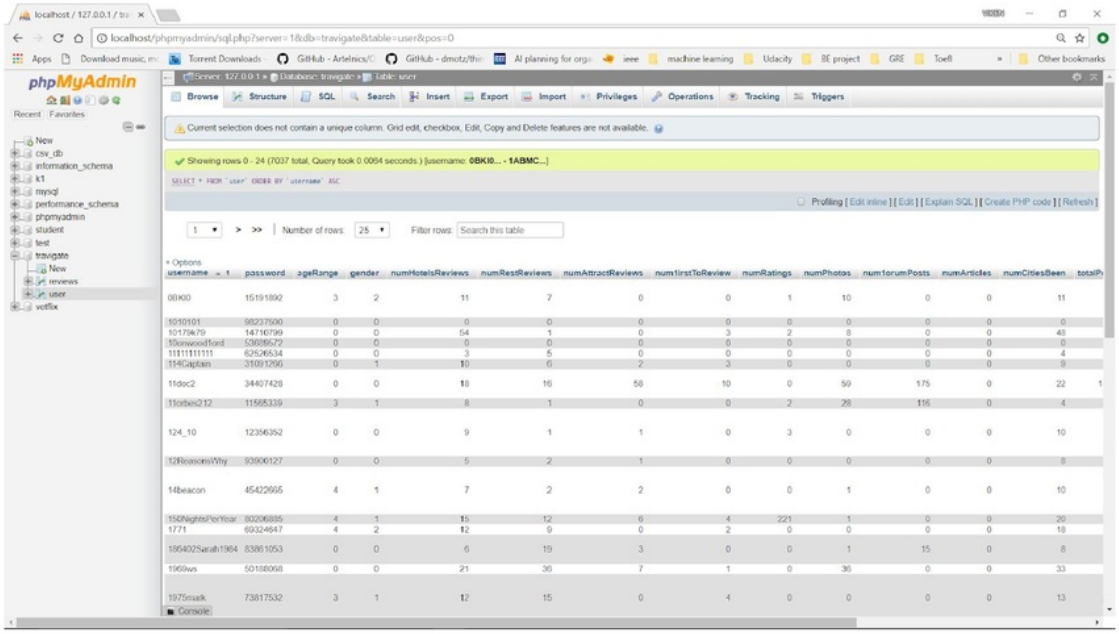


Fig. 4.7: TripAdvisor User Dataset stored in MySQL

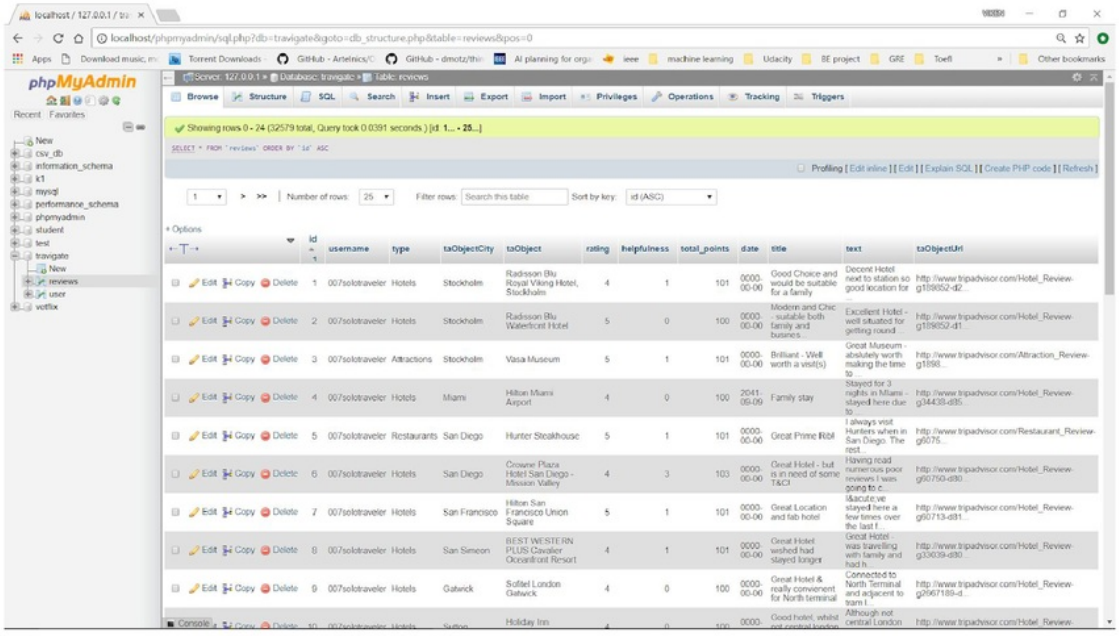


Fig. 4.7: TripAdvisor Review Dataset in MySQL

Chapter 5: TESTING AND RESULTS

In this chapter we discuss various manual test cases and the results/accuracy of the recommendation and recognition modules.

5.1 Testing

1. *User Interface Testing for Client-Server Architecture*

MODULE 1: IMAGE UPLOAD FOR PLACE RECOGNITION				
Test Case ID	**Test Description**	**Test Prerequisite**	**Test Inputs**	**Test Results**
1	Image	An image of a place has to be uploaded	Image of Chhatrapati Shivaji Terminus	Correct placed recognised with 79% accuracy
2	Image	An image of a place has to be uploaded	Image of a puppy	Error message because the image is not of a place
3	Image	An image of a place has to be uploaded	Image of Grand Hyatt Mumbai	Correct placed recognised with 30.3% accuracy
4	Image	An image of a place has to be uploaded	Image of Seawoods Grand Central	Error message because the place is not the training dataset
5	Image	An image of a place has to be uploaded	Image of Chhatrapati Shivaji Terminus	Place detected as The Taj Mahal Palace
6	Image	An image of a place has to be uploaded	Image of Red Fort	Error message because the place is not the training dataset

Table 4: Manual Testing for Uploading the Image in the Place recognition Module

MODULE 2: LOGIN				
Test Case ID	**Test Description**	**Test Prerequisite**	**Test Inputs**	**Test Results**
1	Username	Alphabets	AVSENT	Accepted
2	Username	Alphabets and numbers	A1Mom06	Accepted
3	Username	Alphabets and numbers without special characters	30Brooklyn50#	Rejected
4	Username	Alphabets and numbers without special characters except '_'	AM__26__AP	Accepted
5	Password	Minimum 8 characters	57490208	Accepted
6	Password	Minimum 8 characters	845	Rejected
7	Password	Maximum 20 characters	fdohn6492n	Accepted
8	Password	Maximum 20 characters	gjrolxbndgt589sek,fbn496t7cnfolsbgf	Rejected

Table 5: Manual Testing for Login Module

MODULE 3: NAME OF PLACE				
Test Case Id	**Test Description**	**Test Prerequisite**	**Test Inputs**	**Test Results**
1	Cityname	Alphabets only	Miami	Accepted
2	Cityname	Case sensitive alphabets	DELHI	Accepted
3	Cityname	No numericals allowed	sector 38, Delhi	Rejected
4	Cityname	Special characters other then () not allowed	Mumbai (Bombay)	Accepted
5	Cityname	Special characters other then () not allowed	London!	Rejected
6	Cityname	Only places from the dataset allowed	Vidyavihar	Rejected
7	Cityname	Only places from the dataset allowed	Stockholm	Accepted

Table 6: Manual Testing for Name of Place Module

2. Testing Recommendation and Recognition Modules

- Recommendation System will be tested for 1000 places. The places recommended and the places not recommended will be checked against these pair of similar users and using their reviews about the place, a confusion matrix for the same will be prepared.
- For Recognition System, we divide the initial data that we have from web scraping into 70:20:10 ratio where from the 100 images of each place, 70 images were used for training, 20 will be used for testing and 10 for validation. We have written a tester module to provide us with the accuracy of the system.

5.2 Results

The unsupervised method of K-Modes clustering provides a classification of users into 4 clusters based on their features. The attributes of the 'user' dataset from the TripAdvisor dataset considered for training the clusters are 'age range', 'gender', 'number of hotels reviewed', 'number of first to review', 'number of ratings', 'number of photos', 'number of forum posts', 'number of articles', 'number of cities been', 'total points', 'contributor level', 'number of helpful votes', 'reviewer badge', and 'travel style'. There are entries for about 7000 users and their about 32000 reviews for over 26000 places in the training dataset.

For testing the system, we considered 1000 places with 2 similar users for each place, among which one user is who hasn't reviewed the place as testing data and another user who has reviewed the place to verify it. Similar users are selected based on users that have more than a threshold of similar attributes. Places that have ratings of 3 stars and above by the users who have visited the place are labeled suggested places and rest are not suggested places. Table 2 shows a confusion matrix and hence an accuracy of 89.5%.

	Suggested Places	Not Suggested Places
Suggested Places	249	77
Not Suggested Places	28	646

Table 2: Confusion Matrix of 1000 test places with 2 similar users for each place for the place recommendation system.

The Optimal CNN was developed with 50 steps per epoch for 100 epochs with 200 validation steps to train the model with used for place recognition for the 55 places in Mumbai (Bombay) which provided an optimal average accuracy of 46.4%. Table 3 provides accuracy of image recognition for each of the 55 places trained with 100 images for each place in Mumbai (Bombay) of this optimal CNN model.

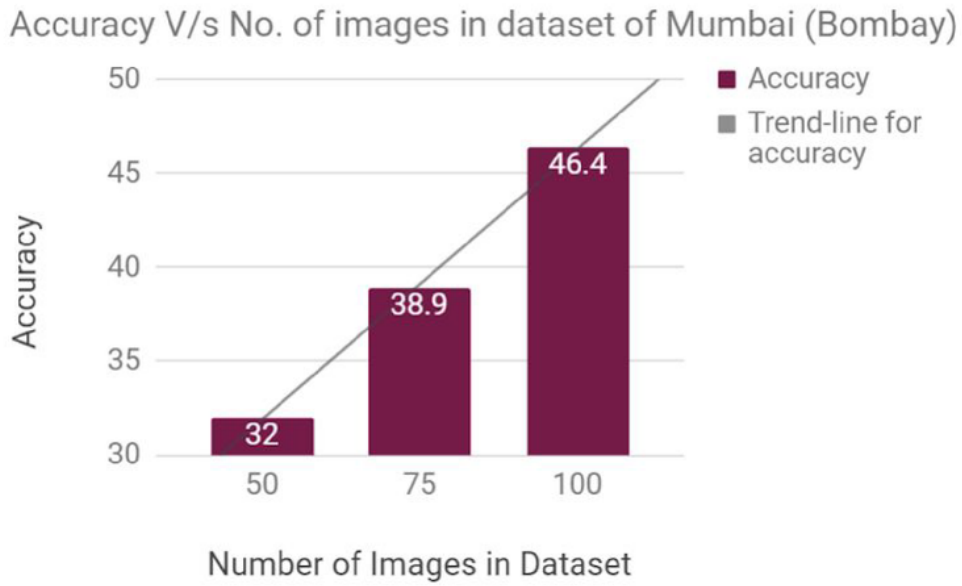


Fig. 5.1: Graph showing Accuracy V/s No. of images in the dataset of Mumbai (Bombay)

Also, as the number of images per place was increased from 50 images per place to 75 images per place, and later to 100 images per place for 55 places in Mumbai (Bombay), the accuracy of CNN model gradually increased as well. The accuracy of the model with 50 images was 32%, accuracy for the model with 75 images was 38.9% and accuracy for the model with 100 images was 46.4%.

The initial CNN with 800 steps per epoch for 25 epochs with 200 validation steps to train the model used for place recognition for the 55 places in Mumbai (Bombay) which provided an accuracy of 35.9%.

No.	Places in Mumbai (Bombay)	Accuracy
1.	21 Fahrenheit Icelounge	51.52%
2.	Ayub's	51.52%
3.	Bademiya	49.49%
4.	bagel shop	55.56%
5.	Bandra-Worli Sea Link	24.24%
6.	Blue Bird Hotel	34.34%
7.	Busaba	34.34%
8.	Cafe Madras	41.41%
9.	Chhatrapati Shivaji Terminus	79.8%
10.	Chings	74.75%
11.	Cowie's Hotel	37.37%
12.	Delhi Darbar	50.51%
13.	Dhobi Ghat	64.65%
14.	Fariyas Hotel Mumbai	58.59%
15.	Flag's	68.69%
16.	Gateway of India	86.87%
17.	Grand Hyatt Mumbai	30.3%
18.	Hotel Airlink	64.65%
19.	Hotel Diplomat	41.41%
20.	Ibis Mumbai Airport	73.74%
21.	InterContinental Marine Drive	22.22%
22.	ITC Grand Central	28.28%
23.	Jewel of India & Jade Garden	43.43%
24.	Kamling	27.27%
25.	Kamran Residency	44.44%
26.	Khyber	54.55%
27.	Lotus Cafe	63.64%
28.	Marine Drive	42.42%
29.	Mumbai Magic - Private Tours	50.51%
30.	Natural Ice Cream Parlour	59.6%
31.	Novotel Mumbai Juhu Beach	45.45%
32.	Out of the Blue	42.42%
33.	Paradise	57.58%
34.	Peshawri	27.27%
35.	Ramee Guestline Hotel, Juhu	32.32%
36.	Renaissance Mumbai Convention Centre Hotel	40.40%
37.	Sahara Star Hotel	50.51%
38.	Sea Palace Hotel	45.45%
39.	Sofitel Mumbai BKC	22.22%
40.	Suba Galaxy	23.23%
41.	Sun-n-Sand Hotel, Mumbai	39.39%
42.	Taj Lands End	49.49%
43.	Thai Pavilion	43.43%
44.	The Fern Residency, Mumbai	53.54%
45.	The Golconda Bowl	60.61%
46.	The LaLiT Mumbai	60.61%
47.	The Leela Mumbai	32.32%
48.	The Mirador Hotel & Restaurant	51.52%
49.	The Taj Mahal Palace	27.27%
50.	The Westin Mumbai Garden City	30.30%
51.	Trident, Bandra Kurla, Mumbai	36.36%
52.	Trident, Nariman Point	67.68%
53.	Vivanta by Taj - President, Mumbai	35.35%
54.	Worli Sea Face	36.36%
55.	Yauatcha	31.31%

Table 3: Accuracy of image recognition of 55 places in Mumbai (Bombay) of a model trained with 100 images per place.

Chapter 6: CONCLUSION AND SCOPE FOR FUTURE WORK

5.1 Conclusion

The central goal of this project was to reduce the efforts of planning a trip. The application proposed by us recommends restaurants, hotels, and attractions in a place that the user (tourist) wants to visit. The app will also recognize a place when given as an image to the system. This system can be specifically useful for the users who are visiting a place for the first time.

We use the K modes algorithm to cluster users with the same interests, these are taken from the training dataset. The results show great accuracy while testing the clusters and suggesting places.

For the place recognition system, we have used the Convolutional Neural Networks, to train and test the system. The results show good accuracy for recognition of places with a high number places to recognize in a city.

The system proposed by us also gives the user an option to add new places from the dataset or review new places. This feature resolves the issue of a limited dataset.

This application can be useful for tourists, locals, and researchers.

5.2 Future Scope

As far as improving the scope is concerned, the accuracy of the system for place recognition can be improved. Several more places across the world can be included if the size of the database or computational power increases, therefore more places can be classified.

Moreover, after every recommendation, or if a user visits a particular place, the system can take feedback from the user and consider the feedback for further recommendations to the same users. Thus a learning system can be created which considers the users feedback after every recommendation.

Bibliography

[1] Zhexue Huang, "Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values" ACSys CRC, CSIRO Mathematical and Information Sciences

[2] Eleonora Pantano, Constantinos-Vasilios Priporas and Nikolaos Stylos, "'You will like it!' using open data to predict tourists' response to a tourist attraction", Elsevier Publication.

[3] Eric Hsueh-Chan Lu, Ching-Yu Chen and Vincent S. Tseng, "Personalized Trip Recommendation with Multiple Constraints by Mining User Check-in Behaviors"

[4] Dominique Lenne, Idir Benouaret, "A Package Recommendation Framework for Trip Planning Activities"

[5] Ana Paula G. Ferreira, Thiago H. Silva and Antonio A. F. Loureiro, "Beyond Sights: Large Scale Study of Tourists' Behavior Using Foursquare Data" 2015 IEEE 15th International Conference on Data Mining Workshops

[6] Juhi Agarwal, Nishkarsh Sharma, Pratik Kumar, Vishesh Parshav, Anubhav Srivastava and R.H. Goudar, "Intelligent Search in E- Tourism Services Using Recommendation System: Perfect Guide for Tourist"

[7] Kyo-Joong Oh, Zaemyung Kim, Hyungrai Oh, Chae-Gyun Lim, and Gahgene Gweon, "Travel Intention-Based Attraction Network for Recommending Travel Destinations"

[8] Xu Chen, Zhiyang Jia, Wei Gao and Yuting Yang, "User-based Collaborative Filtering for Tourist Attraction Recommendations" 2015 IEEE International Conference on Computational Intelligence & Communication Technology.

[9] Neha Sharma and Nirmal Gaud, "K-modes Clustering Algorithm for Categorical Data", International Journal of Computer Applications (0975 – 8887) Volume 127 – No.17, October 2015

[10] Ramsha Fatima, Iffat Zarrin, Mohammed A. Qadeer and M. Sarosh Umar, "Mobile Travel Guide using Image Recognition and GPS/Geo Tagging" 2015 IEEE

[11] Tianyi Liu, Shuangsang Fang, Yuehui Zhao, Peng Wang and Jun Zhang, "Implementation of Training Convolutional Neural Networks.

[12] Ville Hautamaki, Antti Pollanen, Tomi Kinnunen, Kong Aik Lee, Haizhou Li , and Pasi Franti, "A Comparison of Categorical Attribute Data Clustering Methods" School of Computing, University of Eastern Finland, Finland , Institute for Infocomm Research, A*STAR, Singapore.

Appendix A

1. Deployment Details

a. Deploying on Amazon web services (Cloud Computing for Education)

i. Creating MySQL database instance

- Sign in to the AWS Management Console.
- open the Amazon RDS console at https://console.aws.amazon.com/rds/.
- Choose MySQL in the select engine window, and Launch DB instance.

ii. Cleaning and loading data on the MySQL Database instance

- Python script to load data in the excel sheets to MySQL Database instance .
- Load user data from the user dataset in xlsx format.
- Load reviews data from the reviews dataset in xlsx format.

iii. Creating Amazon Elastic Compute 2 Instance

- Open the Amazon Amazon Elastic Compute 2 console at https://console.aws.amazon.com/ec2/ and Launch Instance.
- Choose an Amazon Machine Image (AMI), find an Amazon Ubuntu AMI select it.
- Choose an Instance Type, Configure Instance Details, choose Network, and then choose the entry for your default VPC.
- Choose Subnet, and then choose a subnet in any Availability Zone.
- Add Storage and Launch.
- Install Python 3.6 and it's dependencies on Ubuntu 16.04 using PuTTY.

iv. PuTTY

- PuTTY is a free and open-source terminal emulator which uses SSH to access the Amazon Elastic Compute 2.

- PuTTYgen is used to generate a private key in RSA format and saved with a ".pem" format
- Using the username and public DNS name of the Amazon Elastic Compute 2 instance, "user_name@public_dns_name" and the private key, a putty session is started.
- Run Python scripts in admin mode.

v. WinSCP

- WinSCP is a free and open-source client for Microsoft Windows and is used for secure file transfer between a local and remote Amazon Elastic Compute 2.
- Upload the python scripts from the local Microsoft Windows machine to Amazon Elastic Compute 2 using WinSCP.

b. Creating an Android application and deploying it on an android device

i. Run the following commands to generate the ".apk" file which can be used to install on the android device. The first 3 commands are used only once to generate the platform and unique key for the application.

- ionic cordova platform add ios
- ionic cordova platform add android
- keytool -genkey -v -keystore Travigate.keystore -alias Travigate -keyalg RSA -keysize 2048 -validity 10000
- ionic cordova build android --prod --release
- jarsigner -verify -verbose -certs platforms/android/build/outputs/apk/android-release-unsigned.apk
- jarsigner -verbose -sigalg SHA1withRSA -digestalg SHA1 -keystore Travigate.keystore platforms/android/build/outputs/apk/android-release-unsigned.apk alias_name
- C:\Users\Viken\AppData\Local\Android\sdk\build-tools\25.0.0\zipalign -v 4

platforms/android/build/outputs/apk/android-release-unsigned.apk platforms/android/build/outputs/apk/Travigate.apk

- C:\Users\Viken\AppData\Local\Android\sdk\build-tools\25.0.0\apksigner verify platforms/android/build/outputs/apk/Travigate.apk

2. Minimum System Requirements

The minimum system requirements needed for the system to work are:

Software Requirements

1. Python 3.6 (and Flask Server)
2. Android (Ionic 3 Framework)
3. MySQL Database

Hardware Requirements

1. Smartphone (Android 4.1+) with a camera for image input and an active internet connection

3. User's Manual

The following steps need to be followed while using Travigate:

1. Open the application and register yourself.
2. All the details regarding yourself and your interests are to be inserted. Then click "REGISTER".
3. In case you are an existing user, enter your username and password and click on "SIGN IN".
4. Next, you will be directed to the homepage, where there will be four different options to choose from,
    a. You can change the name of the city
    b. You can look for new attractions, hotels, and restaurants in the chosen city.
    c. You can insert an image of a place, to know the name of the place.

d. You can review any place (Attractions, hotels and, restaurants) and give your ratings.

5. There is an option to view your profile, in which, all your interests, your reviews and other attributes about you, like the no. of photos you have or the no. of reviews you have done are all shown.
6. To get out of the system, go to the menu and click on "LOGOUT".

4. Paper Publication

Viken Parikh, Madhura Keskar, Dhwanil Dharia and Pradnya Gotmare "A Tourist Place Recommendation and Recognition System" International Conference on Inventive Communication and Computational Technologies (ICICCT 2018), April 2018

A Tourist Place Recommendation and Recognition System

Viken Parikh
Computer Engineering Department,
K. J. Somaiya College of Engineering,
Vidyavihar, Mumbai
viken.parikh@somaiya.edu

Madhura Keskar
Computer Engineering Department,
K. J. Somaiya College of Engineering,
Vidyavihar, Mumbai
madhura.k@somaiya.edu

Dhwanil Dharia
Computer Engineering Department,
K. J. Somaiya College of Engineering,
Vidyavihar, Mumbai
dhwanil.d@somaiya.edu

Pradnya Gotmare
Computer Engineering Department,
K. J. Somaiya College of Engineering,
Vidyavihar, Mumbai
pradnyagothmare@somaiya.edu

***Abstract*— Tourism, these days involves mass availability and mass participation in holidays. But many times, a tourist cannot decide which place to visit, or where to stay. In this paper, we propose a mobile application, which will take the user's interest and recommend attractions, restaurants, and hotels. The system is trained using the dataset of TripAdvisor. The clustering of the training dataset is done using K-modes clustering which is an unsupervised learning algorithm. The application Travigate, not only recommends new places to the user, but it also helps them to recognize new places. With the use of Convolutional Neural Networks, reverse image search is done for a dataset created by web scraping images from Google. The application receives the data in the JSON format from the MySQL Database using Python Flask Server.**

***Keywords*—K-Modes Clustering, TripAdvisor, JSON, Collaborative Filtering, Convolutional Neural Networks, Image Recognition, keras, Unsupervised Learning**

I. INTRODUCTION

Tourism has become an important sector that has an impact on the development of the country's economy as on this date. That being said, it is also important to note that tourism and traveling is one of the biggest entertainment in today's modern lifestyle. And the most crucial task, to carry this out efficiently is to plan the trip. Conventionally this planning was done by travel agents [3]. For example, a particular travel company would plan the tours of a particular place, showing the prime tourist attractions. For instance, a company planning a trip to Mumbai (India) would suggest the more common places like Gateway of India or Marine Drive. However, this trip won't satisfy everyone. If a user is interested in places like museums and aquariums, then they would have to plan such a trip on their own.

Nowadays, mobile phones are a necessary part of people's life. There is a continuous rise in the number of mobile computing applications centered on people's daily life. One of the areas in which the user can benefit from smartphone applications is tourism & traveling [10].

We aim at combining the comfort of mobile phones and the need of planning tours by proposing the application of Travigate, wherein, on the basis of similar interests with entries in the training dataset, we recommend places to a user. We use K-modes as a clustering algorithm for categorical data. K-Means is the widely used numerical clustering method where Euclidean distance is used as a distance measure [9]. K-modes is generally used for large datasets and the dataset we have used for this paper is the TripAdvisor dataset. There are entries for about 7000 users and their about 32000 reviews for over 26000 places.

Further, when a tourist visits a place, they are unaware of the place and usually require a tourist guide to tell them which place it is and provide relevant information. Our system has been trained to recognize a place based on image input of the place using the list of places from the TripAdvisor dataset.

II. BASIC TERMINOLOGIES USED

K-modes- Basic Concept

Categorical data clustering is an important research problem in pattern recognition and data mining. Clustering is a

technique used place data items into related groups. These groups, called clusters, have high intra cluster similarity and low inter cluster similarity. K-Means is the widely used numerical clustering method where Euclidean distance is used as a distance measure. Categorical attributes are values with small domains. It is not possible to find the distance between male and female. Hence K-means cannot be used for categorical values. K-modes on the other hand clusters categorical values based on the dissimilarity matrix where the distance is either 0 or 1.

B. CNN Basic Concept

Convolutional Neural Networks (CNN), are a class of deep, feed-forward artificial neural networks. CNN is majorly applied in the fi of Computer Visions. It also has other applications like image and video recognition, recommender systems and natural language processing.

The CNN architecture has three lay . Convolution layer, pooling layer and the output layer. Each convolution layer in the convolution neural network is followed by a computing layer which is used to calculate the local average and the second extract, this unique two feature extraction structure reduces [11].

C. JSON

JSON stands for JavaScript Object Notation. JSON is a language-independent text format that uses conventions of many programming languages making it an ideal data-interchange language. It is built on universal data types supported by all modern programming languages making JSON an ideal data-interchange language [10].

III. RELATED WORK

In the literature, there are several studies about users' l ts and routines in a city using digital traces. Some of them analyzed GPS data and cellular footprints of users to understand, for instance, their usual trajectories.

	Google Goggles	**GuideMe**	**Our Proposed System**
Input	Image	Image/ Text	Image, Text
Output	i) Recognize image ii) Give Name iii) Web Search Results	i) Recognize image ii) Give Name iii) Give Details	i) Recognize image ii) Suggest nearby places in the given city based on users' interests.
Image Category	Any Image	Images of Monuments or famous buildings	Images of 1. Hotels 2. Attractions 3. Restaurants

Table 1: Comparison of various similar systems for place recognition.

A Recommendation System is a personalization tool that offers users with a list of items that best fit their individual taste. A Recommendation System analyses the available data and recommends the items according to user's interest. The more the system understands the interest of the users, the better recommendations, it can perform.

Recommendation System was proposed to cope with the problem of information overload. There are so many traditional methods available for recommending the items to the users. For example, we have content-based filtering, Collaborative based Filtering, and knowledge-based filtering but some of these methods have some shortcomings. In recommendation is totally based on the evaluation in which we will ask the questions to the user and observe the answer [6].

A. Collaborative Filtering:

laborative Filtering is used in recommendation systems. There are two types of collaborative filtering systems, Item-based and Content-based Collaborative Filtering.

In the more general sense, Collaborative Filtering is the process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc. There are various reasons why Collaborative Filtering is used like it can be used with large datasets. As compared to content-based filtering, the accuracy of Collaborative Filtering is much more.

Using Collaborative Filtering the whole process of recommending tourists places can be divided into three steps:

a. The representation of tourist information, wherein the travel s e and reviews of the user are checked.
b. Next, the similarity of tourists can be computed according to the visiting history data and the Collaborative Filtering algorithm presented by us.
c. The generation of attraction recommendations. On the basis of similarities with other similar users, the top Attractions, Restaurants, and Hotels are recommended.

B. Recommender System:

A recommender system or a recommendation system is a subclass of information filtering system that seeks to predict the rating or preference that a user would give to an item.

Classical recommender systems provide the user with recommendations as ranked lists consisting of single items,

e.g., movie, book. In trip planning, a user is interested in suggestions for points of interest (POI), that could be very heterogeneous, e.g., museum, park, restaurant, etc. A tourism recommender system can benefit from a system capable of recommending items organized in packages (bundles) rather than ranked lists, which constitute an improved exploration experience for the visitor [7].

IV. DATASET DESCRIPTION

Two different datasets have been used. For users' profiles and places' reviews database, the Tripadvisor dataset has been used. For image recognition which consists of images crawled from Google.

A. *TripAdvisor Dataset*

There are entries for about 7000 users and their about 32000 reviews for over 26000 places across the world.

a. *TripAdvisor User Dataset:*

Each reviewer can choose a travel style among 17 different categories (Multiple options allowed) such as Foodie, Luxury Traveler, Shopping Fanatic, Urban Explorer, Peace and Quiet Seeker, Art and Architecture Lover, Thrifty Traveler, Vegetarian, History Buff, Thrill Seeker, Beach Goer, Family Holiday Maker, Nature Lover, Trendsetter, Like a Local, Nightlife Seeker, 60+ Traveler. There are also several other entries such as age range, Number of Helpful reviews, total reviews, number of cities been, contribution level, number of helpful votes etc.

b. *TripAdvisor Reviews Dataset:*

There are mainly 3 categories of places (Hotels, Restaurants, and Attractions). Each tuple has the following entries- username, type, date, title, text, rating, helpfulness, total-points, Name of Place, City (where the place is located).

B. *Images Dataset*

The images dataset has been crawled from Google Images for the places mentioned in the TripAdvisor dataset.

For our training model 55 places that have been reviewed in Mumbai (Bombay) from the TripAdvisor dataset with 100 images for each place which has been achieved using our web crawler designed for scrapping first 100 images from google images search. We split this data into 70 images training, 20 images validation and 10 images for testing each class in the dataset.

V. METHODOLOGY

A. *Place Recommendation System*

The system consists of a MySQL database, a Web Server, and a front-end application.

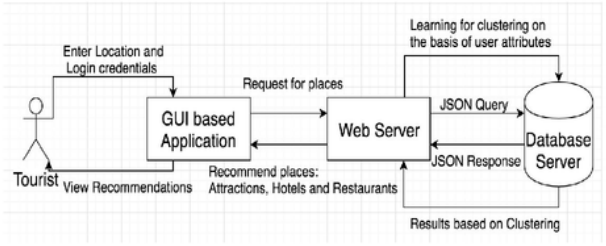


Fig. 1: Recommendation system architecture

a. *Data Cleaning and Data Pre-Processing:*

The initial step the data was initially in the form of .csv files it had to be converted into a MySQL database. The dataset also required cleaning and removing redundant or missing data from the datasets. Moreover, it was also necessary converting all the data into categorical data.

b. *Recommendation Using K-Modes:*

The users are clustered using K-modes clustering algorithm, where the total number of clusters can be n (n>2). As the number of clusters increases, the number of users in each cluster decrease. Based on the clusters, if a user wants recommendations for places in a particular city, the system would recommend places users have visited in the same cluster.

For example, if the user wants recommendations, for places to visit in Mumbai (Bombay). The system would recommend places in Mumbai (Bombay) which have been visited by users in the same cluster.

$$d(X, Y) = \sum_{j=1}^{m} \delta(x_j, y_j) \quad (1)$$

where,

$$\delta(x_j, y_j) = \begin{cases} 0, & x_j = y_j \\ 1, & x_j \neq y_j \end{cases} \quad (2)$$

Where d(X, Y) is the distance function, to find the dissimilarity between X and Y described by m categorical values.

B. *Place Recognition System*

a. *Image Scraping:*

To create a sufficiently large training dataset for different places, we used the web scraping technique, wherein the first hundred images of a Google search of that particular place are scraped.

In the image search page created by Google, the first hundred websites are visited and the images from those websites are downloaded and stored in individual folders.

This method is convenient when we do not have an existing dataset. We have attempted to use this technique to acquire many photos of a particular place, in order to train the system.

b. Initializing CNN:

There are a number of important choices that need to be made when we are building and training a neural network. We had to decide how many layers to have, what stride and kernel size to use for each convolution layer and many more. But amongst all these, weight initialization has the greatest impact on both the convergence rate and final quality of a network.

Thus, in this part of the implementation, we initialize the weights used in this Convolutional Neural Networks.

c. Convolution:

After procuring these images and initializing the weights, the next step is the application of the convolution layer. Convolution is basically a mathematical operation which provides a way of multiplying two arrays of numbers, generally of different sizes, but of the same dimensionality, to produce a third array of numbers of the same dimensionality.

In our code, we have used the method Conv2D(), which is using a 2D convolution layer, this layer creates a convolution kernel that is convolved with the layer input to produce a tensor of outputs.

d. Pooling:

A pooling layer is periodically inserted in between two convolution layers. This layer, is used here to downsample the volume spatially, independently in each depth slice. The method we have used is that of MaxPooling2D(). This is used for the maximum pooling, between two convolution layers.

e. Flattening:

After another convolution layer and a max pooling layer which reduces height and width of the individual channels in favor of their number, i.e., depth, we use the flattening layer.

For the application of the last layer, the dense layer, it is necessary to convert the output of the convolutional part into a 1D array. This is achieved by flattening the output of the initial layers.

f. Dense Layer:

The dense layer is the third layer in the Convolutional Neural Networks.
In a dense layer, all the nodes are connected i.e., every node in the layer is connected to every node in the preceding layer. A dense layer thus is used to change the dimensions of your vector.

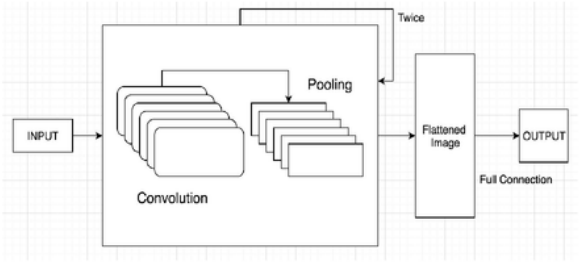


Fig. 2: Architecture of CNN algorithm

VI. RESULTS

The unsupervised method of K-Modes clustering provides a classification of users into 4 clusters based on their features. The attributes of the 'user' dataset from the TripAdvisor dataset considered for training the clusters are 'age range', 'gender', 'number of hotels reviewed', 'number of first to review', 'number of ratings', 'number of photos', 'number of forum posts', 'number of articles', 'number of cities been', 'total points', 'contributor level', 'number of helpful votes', 'reviewer badge', and 'travel style'. There are entries for about 7000 users and their about 32000 reviews for over 26000 places in the training dataset.

For testing the system, we considered 1000 places with 2 similar users for each place, among which one user is who hasn't reviewed the place as testing data and another user who has reviewed the place to verify it. Similar users are selected based on users that have more than a threshold of similar attributes. Places that have ratings of 3 stars and above by the users who have visited the place are labeled suggested places and rest are not suggested places. Table 2 shows a confusion matrix and hence an accuracy of 89.5%.

	Suggested Places	Not Suggested Places
Suggested Places	249	77
Not Suggested Places	28	646

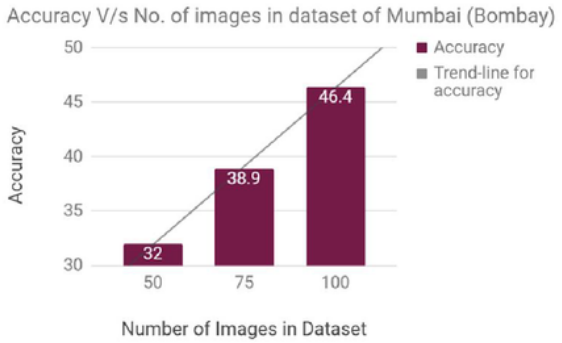
Table 2: Confusion Matrix of 1000 test places with 2 similar users for each place for the place recommendation system.

The initial CNN with 800 steps per epoch for 25 epochs with 200 validation steps to train the model used for place recognition for the 55 places in Mumbai (Bombay) which provided an accuracy of 35.9%.

No.	Places in Mumbai (Bombay)	Accuracy
1.	21 Fahrenheit Icelounge	51.52%
2.	Ayub's	51.52%
3.	Bademiya	49.49%
4.	bagel shop	55.56%
5.	Bandra-Worli Sea Link	24.24%
6.	Blue Bird Hotel	34.34%
7.	Busaba	34.34%
8.	Cafe Madras	41.41%
9.	Chhatrapati Shivaji Terminus	79.8%
10.	Chings	74.75%
11.	Cowie's Hotel	37.37%
12.	Delhi Darbar	50.51%
13.	Dhobi Ghat	64.65%
14.	Fariyas Hotel Mumbai	58.59%
15.	Flag's	68.69%
16.	Gateway of India	86.87%
17.	Grand Hyatt Mumbai	30.3%
18.	Hotel Airlink	64.65%
19.	Hotel Diplomat	41.41%
20.	Ibis Mumbai Airport	73.74%
21.	InterContinental Marine Drive	22.22%
22.	ITC Grand Central	28.28%
23.	Jewel of India & Jade Garden	43.43%
24.	Kamling	27.27%
25.	Kamran Residency	44.44%
26.	Khyber	54.55%
27.	Lotus Cafe	63.64%
28.	Marine Drive	42.42%
29.	Mumbai Magic - Private Tours	50.51%
30.	Natural Ice Cream Parlour	59.6%
31.	Novotel Mumbai Juhu Beach	45.45%
32.	Out of the Blue	42.42%
33.	Paradise	57.58%
34.	Peshawri	27.27%
35.	Ramee Guestline Hotel, Juhu	32.32%
36.	Renaissance Mumbai Convention Centre Hotel	40.40%
37.	Sahara Star Hotel	50.51%
38.	Sea Palace Hotel	45.45%
39.	Sofitel Mumbai BKC	22.22%
40.	Suba Galaxy	23.23%
41.	Sun-n-Sand Hotel, Mumbai	39.39%
42.	Taj Lands End	49.49%
43.	Thai Pavilion	43.43%
44.	The Fern Residency, Mumbai	53.54%
45.	The Golconda Bowl	60.61%
46.	The LaLiT Mumbai	60.61%
47.	The Leela Mumbai	32.32%
48.	The Mirador Hotel & Restaurant	51.52%
49.	The Taj Mahal Palace	27.27%
50.	The Westin Mumbai Garden City	30.30%
51.	Trident, Bandra Kurla, Mumbai	36.36%
52.	Trident, Nariman Point	67.68%
53.	Vivanta by Taj - President, Mumbai	35.35%
54.	Worli Sea Face	36.36%
55.	Yauatcha	31.31%

Table 3: Accuracy of image recognition of 55 places in Mumbai (Bombay) of a model trained with 100 images per place.

The Optimal CNN was developed with 50 steps per epoch for 100 epochs with 200 validation steps to train the model with used for place recognition for the 55 places in Mumbai (Bombay) which provided an optimal average accuracy of 46.4%. Table 3 provides accuracy of image recognition for each of the 55 places trained with 100 images for each place in Mumbai (Bombay) of this optimal CNN model.



Graph 1: Accuracy V/s No. of images in the dataset of Mumbai (Bombay)

Also, as the number of images per place were increased from 50 images per place to 75 images per place, and later to 100 images per place for 55 places in Mumbai (Bombay), the accuracy of CNN model gradually increased as well. The accuracy for the model with 50 images was 32%, accuracy for the model with 75 images was 38.9% and accuracy for the model with 100 images was 46.4%.

VII. CONCLUSION

The central goal of this project was to reduce the efforts of planning a trip. The application proposed by us recommends restaurants, hotels, and attractions in a place that the user (tourist) wants to visit. The app will also recognize a place when given as an image to the system. This system can be specifically useful for the users who are visiting a place for the first time.

We use the K modes algorithm to cluster users with the same interests, these are taken from the training dataset. The results show great accuracy while testing the clusters and suggesting places.

For the place recognition system, we have used the Convolutional Neural Networks, to train and test the system.

The results show good accuracy for recognition of places with a high number places to recognize in a city.

This application can be useful for tourists, locals, and researchers.

VIII. FUTURE WORK

The accuracy of the system for place recognition can be improved. Several more places across the world can be included if the size of the database or computational power increases, therefore more places can be classified. Moreover, after every recommendation, or if a user visits a particular place, the system can take feedback from the user and consider the feedback for further recommendations to the same users. Thus a learning system can be created which considers the users feedback after every recommendation.

IX. REFERENCES

[1] Zhexue Huang, "Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values" ACSys CRC, CSIRO Mathematical and Information Sciences

[2] Eleonora Pantano, Constantinos-Vasilios Priporas and Nikolaos Stylos, "'You will like it!' using open data to predict tourists' response to a tourist attraction", Elsevier Publication.

[3] Eric Hsueh-Chan Lu, Ching-Yu Chen and Vincent S. Tseng, "Personalized Trip Recommendation with Multiple Constraints by Mining User Check-in Behaviors"

[4] Dominique Lenne, Idir Benouaret, "A Package Recommendation ework for Trip Planning Activities"

[5] Ana Paula G. Ferreira, Thiago H. Silva and Antonio A. F. Loureiro, "Beyond Sights: Large Scale Study of Tourists' Behavior Using Foursquare Data" 2015 IEEE 15th International Conference on Data ning Workshops

[6] Juhi Agarwal, Nishkarsh Sharma, Pratik Kumar, Vishesh Parshav, Anubhav Srivastava and R.H. Goudar, "Intelligent Search in E-Tourism Services Using Recommendation System: Perfect Guide for Tourist"

[7] Kyo-Joong Oh, Zaemyung Kim, Hyungrai Oh, Chae-Gyun Lim, and Gahgene Gweon, "Travel Intention-Based Attraction Network for Recommen Travel Destinations"

[8] Xu Chen, Zhiyang Jia, Wei Gao and Yuting Yang, "User-based Collaborative Filtering for Tourist Attraction Recommendations" 2015 IEEE International Conference on Computational Intelligence & mmunication Technology.

[9] Neha Sharma and Nirmal Gaud, "K-modes Clustering Algorithm for tegorical Data", International Journal of Computer Applications 975 – 8887) Volume 127 – No.17, October 2015

[10] Ramsha Fatima, Iffat Zarrin, Mohammed A. Qadeer and M. Sarosh Umar, "Mobile Travel Guide using Image Recognition and GPS/Geo ging" 2015 IEEE

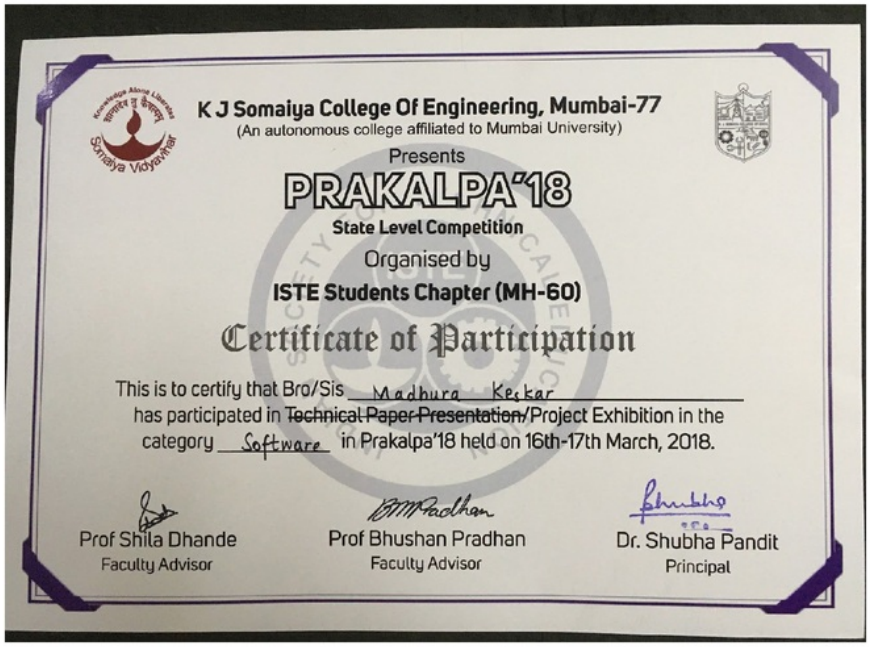
[11] Tianyi Liu, Shuangsang Fang, Yuehui Zhao, Peng Wang and Jun Zhang, "Implementation of Training Convolutional Neural Networks"

Certificates

K J Somaiya College Of Engineering, Mumbai-77
(An autonomous college affiliated to Mumbai University)
Presents
PRAKALPA'18
State Level Competition
Organised by
ISTE Students Chapter (MH-60)
Certificate of Participation

This is to certify that Bro/Sis Dhwanil Dharia
has participated in ~~Technical Paper Presentation~~/Project Exhibition in the category Software in Prakalpa'18 held on 16th-17th March, 2018.

Prof Shila Dhande, Faculty Advisor
Prof Bhushan Pradhan, Faculty Advisor
Dr. Shubha Pandit, Principal



K J Somaiya College Of Engineering, Mumbai-77
(An autonomous college affiliated to Mumbai University)
Presents
PRAKALPA'18
State Level Competition
Organised by
ISTE Students Chapter (MH-60)
Certificate of Participation

This is to certify that Bro/Sis Madhura Keskar
has participated in ~~Technical Paper Presentation~~/Project Exhibition in the category Software in Prakalpa'18 held on 16th-17th March, 2018.

Prof Shila Dhande, Faculty Advisor
Prof Bhushan Pradhan, Faculty Advisor
Dr. Shubha Pandit, Principal

K J Somaiya College Of Engineering, Mumbai-77

(An autonomous college affiliated to Mumbai University)

Presents

PRAKALPA'18

State Level Competition

Organised by

ISTE Students Chapter (MH-60)

Certificate of Participation

This is to certify that Bro/Sis Viken Parikh has participated in ~~Technical Paper Presentation~~/Project Exhibition in the category Software in Prakalpa'18 held on 16th-17th March, 2018.

Prof Shila Dhande
Faculty Advisor

Prof Bhushan Pradhan
Faculty Advisor

Dr. Shubha Pandit
Principal

IEEE

Certificate of Participation

This is to certify that

Dhwanil Dhasia

has successfully presented a paper entitled

A Tourist Place Recommendation and Recognition System

in the International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) during 20-21 April 2018.

Session Chair **Organizing Secretary** **Conference Chair**

ICICCT-2018

IEEE

Certificate of Participation

This is to certify that

Madhura Kaskar

has successfully presented a paper entitled

A Tourist Place Recommendation and Recognition System

in the International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) during 20-21 April 2018.

Session Chair **Organizing Secretary** **Conference Chair**

ICICCT-2018

IEEE

Certificate of Participation

This is to certify that

Viken Parikh

has successfully presented a paper entitled

A Tourist Place Recommendation and Recognition System

in the International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) during 20-21 April 2018.

Session Chair **Organizing Secretary** **Conference Chair**

ICICCT-2018

Plagiarism Report

Acknowledgements

The completion of our project has valuable contribution from various people with whom we have interacted within the organization. This study, which is a complete fulfillment of the requirement for the degree of Bachelor of Technology, would not have been possible without the constant and support, mutual coordination of several people to whom we owe our sincere gratitude.

We would first and foremost like to thank our advisor, Prof. Pradnya Gotmare (Associate Head of Department at K. J. Somaiya College of Engineering, Mumbai) for providing insights and helping us throughout. We feel obliged to ma'am for her kind supervision and valuable guidance which has got our project to this level.

We would also like to thank our Head Of Department Prof. Jyoti Rao, who extended every facility to us for making and completing this project smoothly. We would also like to thank our fellow peers and friends who constantly challenge us while also inspiring us, and provide us with their selfless help and invaluable advices.

We thank the teaching and non – teaching staff of our college as well as the library staff for providing us timely essential information in form of books and for their assistance in the laboratories.

Lastly, we would like to deeply acknowledge our parents for their constant care and support throughout the duration of our project.

A Tourist Place Recommendation and Recognition System - A Plagiarism Report

ORIGINALITY REPORT

20%	14%	15%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Ramsha Fatima, Iffat Zarrin, Mohammed A. Qadeer, M. Sarosh Umar. "Mobile travel guide using image recognition and GPS/Geo tagging: A smart way to travel", 2016 Thirteenth International Conference on Wireless and Optical Communications Networks (WOCN), 2016 Publication	2%
2	gigapaper.ir Internet Source	2%
3	www.ijcaonline.org Internet Source	1%
4	Submitted to Colorado Technical University Online Student Paper	1%
5	dblp.dagstuhl.de Internet Source	1%
6	Idir Benouaret, Dominique Lenne. "A Package Recommendation Framework for Trip Planning	1%

Activities", Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16, 2016
Publication

7	Submitted to VIT University Student Paper	1%
8	Vaishnavi Bheemarao Joshi, R H Goudar. "Searching, categorizing and tour planning: A novel approach towards e-tourism", 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2017 Publication	1%
9	Zhiyang Jia, Yuting Yang, Wei Gao, Xu Chen. "User-Based Collaborative Filtering for Tourist Attraction Recommendations", 2015 IEEE International Conference on Computational Intelligence & Communication Technology, 2015 Publication	<1%
10	arxiv.org Internet Source	<1%
11	Submitted to Beykent Universitesi Student Paper	<1%
12	bristol.ac.uk Internet Source	<1%

13	etds.lib.ncku.edu.tw Internet Source	<1%
14	P.A. Riyaz, Surekha Mariam Varghese. "A Scalable Product Recommendations Using Collaborative Filtering in Hadoop for Bigdata", Procedia Technology, 2016 Publication	<1%
15	S. Aranganayagi, K. Thangavel, S. Sujatha. "New distance measure based on the domain for categorical data", 2009 First International Conference on Advanced Computing, 2009 Publication	<1%
16	www.automationlearn.com Internet Source	<1%
17	Submitted to Institute of Technology Blanchardstown Student Paper	<1%
18	en.wikipedia.org Internet Source	<1%
19	Ana Paula G. Ferreira, Thiago Henrique Silva, Antonio Alfredo Ferreira Loureiro. "Beyond Sights: Large Scale Study of Tourists' Behavior Using Foursquare Data", 2015 IEEE International Conference on Data Mining Workshop (ICDMW), 2015 Publication	<1%

20	keras.io Internet Source	<1%
21	www.tripadvisor.com Internet Source	<1%
22	Submitted to University of Sri Jayewardenepura Nugegoda Sri Lanka Student Paper	<1%
23	Submitted to October University for Modern Sciences and Arts (MSA) Student Paper	<1%
24	Submitted to Indian School of Business Student Paper	<1%
25	www.opengardensblog.futuretext.com Internet Source	<1%
26	Han Mei, Xiangmin Xu. "EEG-based emotion classification using convolutional neural network", 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), 2017 Publication	<1%
27	Submitted to UNITEC Institute of Technology Student Paper	<1%
28	www.gabormelli.com Internet Source	<1%
	www.di.unito.it	

29	Internet Source	<1%
30	docs.aws.amazon.com Internet Source	<1%
31	ieeexplore.ieee.org Internet Source	<1%
32	Submitted to University of Warwick Student Paper	<1%
33	dl.acm.org Internet Source	<1%
34	Lecture Notes in Computer Science, 2014. Publication	<1%
35	Mehdi Ghayoumi, Arvind K Bansal. "Multimodal architecture for emotion in robots using deep learning", 2016 Future Technologies Conference (FTC), 2016 Publication	<1%
36	www.tensorflow.org Internet Source	<1%
37	www.icmars.org Internet Source	<1%
38	Submitted to Bolton Institute of Higher Education Student Paper	<1%

39	bountify.co Internet Source	<1%
40	Submitted to Midlands State University Student Paper	<1%
41	www.scribd.com Internet Source	<1%
42	Jensen. "Classification Methods", Computational Intelligence and Feature Selection, 09/09/2008 Publication	<1%
43	Submitted to Western Governors University Student Paper	<1%
44	www.mysciencework.com Internet Source	<1%
45	annals-csis.org Internet Source	<1%
46	sueboy.blogspot.com Internet Source	<1%
47	Joshua Huang. "Clustering of SNP Data with Application to Genomics", Sixth IEEE International Conference on Data Mining - Workshops (ICDMW 06), 12/2006 Publication	<1%
48	"Biologically Rationalized Computing	<1%

Techniques For Image Processing Applications", Springer Nature, 2018
Publication

49	Submitted to King Saud University Student Paper	<1%
50	Submitted to Sandra Day O'Connor High School Student Paper	<1%
51	"Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012", Springer Nature, 2014 Publication	<1%
52	Chenyi Zhang, Hongwei Liang, Ke Wang. "Trip Recommendation Meets Real-World Constraints", ACM Transactions on Information Systems, 2016 Publication	<1%
53	Eleonora Pantano, Constantinos-Vasilios Priporas, Nikolaos Stylos. "'You will like it!' using open data to predict tourists' response to a tourist attraction", Tourism Management, 2017 Publication	<1%
54	K. Sundaramoorthy, R. Durga, S. Nagadarshini. "NewsOne — An Aggregation System for News Using Web Scraping Method", 2017	<1%

International Conference on Technical Advancements in Computers and Communications (ICTACC), 2017
Publication

55	Submitted to Coventry University Student Paper	<1%
56	Submitted to CSU, San Jose State University Student Paper	<1%
57	www.devinterface.com Internet Source	<1%
58	Yunpeng Liu, Shaotong Pei, Weiping Fu, Kaiyuan Zhang, Xinxin Ji, Zihui Yin. "The discrimination method as applied to a deteriorated porcelain insulator used in transmission lines on the basis of a convolution neural network", IEEE Transactions on Dielectrics and Electrical Insulation, 2017 Publication	<1%
59	Submitted to Aston University Student Paper	<1%
60	Submitted to Texas A & M University, Kingville Student Paper	<1%
61	vigor.co.za Internet Source	<1%
62	David J. Attokaren, Ian G. Fernandes, A.	<1%

Sriram, Y. V. Srinivasa Murthy, Shashidhar G. Koolagudi. "Food classification from images using convolutional neural networks", TENCON 2017 - 2017 IEEE Region 10 Conference, 2017
Publication

---

63 Recommender Systems Handbook, 2011. <1%
Publication

---

64 Submitted to University of Illinois at Urbana-Champaign <1%
Student Paper

---

65 Submitted to Sri Lanka Institute of Information Technology <1%
Student Paper

---

66 Burcu Kantarci, Vincent Labatut. "Classification of Complex Networks Based on Topological Properties", 2013 International Conference on Cloud and Green Computing, 2013 <1%
Publication

---

67 www.buzzparas.com <1%
Internet Source

---

68 Suvarna Bhise, Uday Pandit Khot. "Rayleigh and Ricean Fading Channel simulator-a review", 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2016 <1%
Publication

69	danpritchard.com Internet Source	<1%
70	"Data Science", Springer Nature, 2017 Publication	<1%
71	"Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence", Springer Nature, 2007 Publication	<1%
72	www.sundaychennai.com Internet Source	<1%
73	wwwis.win.tue.nl Internet Source	<1%
74	Dennis Fitzpatrick. "AC Analysis", Elsevier BV, 2018 Publication	<1%
75	eric-yuan.me Internet Source	<1%
76	ijarcce.com Internet Source	<1%
77	Yang, M.-S.. "A fuzzy k-partitions model for categorical data and its comparison to the GoM model", Fuzzy Sets and Systems, 20080216 Publication	<1%
78	Shiva Jalali, Sayyed Alireza Hashemi Golpayegani, Hoda Ghavamipoor. "Designing a	<1%

model of decision making in layers of supply, manufacturing, and distribution of the supply chain: A recommender-based system", 8th International Conference on e-Commerce in Developing Countries: With Focus on e-Trust, 2014

Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

A Tourist Place Recommendation and Recognition System - A Plagiarism Report

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34

PAGE 35

PAGE 36

PAGE 37

PAGE 38

PAGE 39

PAGE 40

PAGE 41

PAGE 42

PAGE 43

PAGE 44

PAGE 45

PAGE 46

PAGE 47

PAGE 48

PAGE 49

PAGE 50

PAGE 51

PAGE 52

PAGE 53

PAGE 54

PAGE 55

PAGE 56

PAGE 57

PAGE 58

PAGE 59

PAGE 60

PAGE 61

PAGE 62

PAGE 63

PAGE 64